

Taking a stand on animal-rights violence

Governments must not turn a blind eye to intimidation and violence by animal-rights activists. A more resilient approach is needed.

Two months ago, the New York Stock Exchange (NYSE) was set to list the shares of Huntingdon Life Sciences (HLS), a British research company whose bankers and advisers have been constantly hounded by animal-rights activists. When the NYSE announced at the last minute that the listing had been shelved, HLS executives weren't the only ones to be dismayed.

At a hearing of the US Senate Committee on Environment and Public Works last week, the NYSE failed to shed much light on the reasons for the withdrawal. Senators James Inhofe (Republican, Oklahoma) and Frank Lautenberg (Democrat, New Jersey) expressed exasperation that a major national institution could leave itself exposed to allegations that it folded in the face of intimidation.

The committee is supporting legislation that would close several loopholes in an existing law that was designed to protect animal researchers. The changes would, among other things, make it easier to prosecute those who encourage violent attacks on employees of companies that don't do animal research themselves, but have business links with firms, such as HLS, that do.

This legislative approach is likely to prove more fruitful than the aggressive pursuit of precisely what happened when the NYSE pulled the plug on the listing. Inhofe had previously written to the exchange and begged it to show some moral backbone. "It seems to me unimaginable that this country's worldwide symbol of the integrity of the capital markets, the NYSE, would capitulate to threats, or even the mere threat of threats, from a single-issue extremist group," the senator said last week. "Appeasing these groups only validates the effectiveness of their tactics and inspires them to replicate this model of activism in some other venue."

Few would disagree with that, and it is appropriate that the Senate committee should try to shed light on a decision that has received very little play in the US press.

HLS is understandably frustrated. Its lawyer, Mark Bibi, described the NYSE's decision as "perhaps the most shameful apparent capitulation to date". He added that the company has received no feedback from the NYSE about the reasons for it — which the NYSE

says are confidential — or about any information that HLS could now provide to help it secure a listing.

But the inclination to blame NYSE and its president, Catherine Kinney — who withdrew from testifying before the Senate, letting her top lawyer take the heat instead — should be resisted. Like other clients and advisers who have shunned HLS, in both Britain and the United States, the NYSE is in a tough position, having had its own staff threatened on the Internet and elsewhere.

Even so, the public mood in the United States could support a more resilient stance by the NYSE and other institutions. There is less latent sympathy for animal-rights activists than in Britain, and more appetite for a vigorous clampdown on their criminal activities. Inhofe and Lautenberg should be applauded for their robust public stance on this issue. It contrasts with an occasional tendency in the United States to wish away hundreds of documented instances of animal-rights-related violence.

But the issue isn't going away. The intellectual leadership of the animal-rights movement is shifting to the United States, and the committee took testimony from Jerry Vlasak, a California-based surgeon, who declined to repudiate previous statements defending violence, and even murder, against researchers. Again, the senators did well to confront the thinking behind animal-rights violence. The committee is now pursuing legal remedies. There is some public fatigue with new laws specific to the motivation of particular crimes, but in this case such laws are needed. Animal-rights activists are exploiting loopholes that, for example, prevent the use of extortion law unless the extorter seeks personal gain. British laws specifically designed to protect animal research were introduced this summer and have had a positive effect. Scientists and national institutions must stand united against animal-rights violence, and legislators should support them by passing Inhofe's proposal. ■

"In the United States there is less latent sympathy for animal-rights activists than in Britain, and more appetite for a clampdown on their criminal activities."

Turkey's evolution

Admission to the European Union can benefit Turkish science.

Turkey is engaged in negotiations for membership of the European Union (EU), and the first such talks, which opened on 18 October, were centred on science and technology. But they took place at a time when many Turkish scientists are at loggerheads with their government, led by the mildly Islamic Justice and Development

Party. They say the government is, by stealth, allowing Islamic influences to infiltrate the constitutionally secular academic system.

When the Turkish Republic was founded by national hero Kemal Atatürk in 1923, it could boast only a few dozen trained physicians and engineers. Its citizens were dirt-poor, and education available to but a few. This legacy of the sultan-caliphs was put into sharp reverse by Atatürk. His modernization programme was unmistakably Western, and could almost have been conceived with membership of the EU in mind.

He changed the alphabet to Latin script that would be readable by Europeans. He introduced education for all, forcing the literacy rate



up from less than 10% to 33% within 15 years. Now 86.5% of Turks are literate. He also abolished the wearing of the veil by women (but not the headscarf), and introduced a constitution solidly anchored in secularism.

At Turkey's western edge, it borders the EU; at the east it borders Iran. As religiosity has grown in Iran since the 1979 Islamic revolution, political tensions in Turkey have grown too. While pragmatically aiming for EU membership, Turkey has also had to deal with the rising confidence of Islamic groups and their growing numbers.

The academic élite — proud adherents to Atatürk's vision — fear this confidence, and their response has been defensive. When headscarves became more common in the 1980s, the Council of Higher Education banned the wearing of them in universities. As the number of special secondary schools for training imams (religious leaders) grew, the council raised the university entrance qualification requirements for students attending these schools above those for normal state schools. The storm over the arrest of the rector of the 100th Year University in Van (see page 8) reflects the bitterness of the struggle within universities to keep Islamic influence at bay.

The academic élite also resents recent government interference in academic appointments. Since his election in 2003, Prime Minister Recep Tayyip Erdogan has passed two contentious laws that affect universities. One allows the government to appoint members of the

board of TÜBİTAK, Turkey's main research agency, which is a major player in the current EU talks. Critics say that subsequent appointments have been politically inspired, and charge that aspects of the agency's current set-up are unconstitutional. A second law requires government approval of university appointments. The government says this is aimed at ending cronyism in the academic world, but critics fear that it will damage academic freedom.

Given this delicate situation, the opening of negotiations for EU membership offers the best hope for the continuing development of science in Turkey. Turkish scientists have little choice but to place their trust in these negotiations.

The government has, to its credit, doubled the science budget in anticipation of the EU talks, and it already pays for Turks to take part in EU Framework programmes as equal partners. Under the watchful eye of EU negotiators, Turkish science will have to be seen to be open, competitive and democratic.

The negotiations will no doubt be protracted, but if they are successful, science in Turkey will be a winner — and part of Atatürk's dream will also have won through. ■

"The opening of negotiations for EU membership offers the best hope for the continuing development of science in Turkey."

Clamp down on copycats

Plagiarism is on the rise, thanks to the Internet. Universities and journals need to take action.

Just how prevalent is plagiarism? At a meeting devoted to the topic at New York University last month, Alan Price of the Office of Research Integrity (ORI), which primarily handles complaints in biomedicine, reported that in the past 16 years, only 5–12% of its misconduct cases each year involved plagiarism. This is defined by the ORI as "the appropriation of another person's ideas, processes, results, or words without giving appropriate credit".

On the other hand, James Kroll, head of administrative investigations at the US National Science Foundation, revealed that more than 60% of its misconduct findings concern plagiarism. And earlier this year, the National Natural Science Foundation of China reported that plagiarism accounted for about one-third of its misconduct cases in the past six years.

Human nature hasn't changed recently, but reusing with the intent to deceive seems to be on the rise, both in the literature and in grant proposals. The replacement of pen and paper with software makes it far easier to slip in large sections of text. Internet connectivity, online repositories and sophisticated search tools provide almost irresistible accessibility to the polished thoughts of others.

Students trained today have grown up in an environment where access is taken for granted and attribution only loosely enforced. So they need more rigorous instruction than their predecessors regarding the ethical standards expected of them. Mentors must counter the ever-rising promotion and funding pressures that reward prolific publication rather than support creative quests.

Although the development of web-based tools that can recognize text-based plagiarism will eventually help detection, more can be done before that point. Some common-sense guidelines need stressing at the bench, long before the data or grant application are written up. Copying text, even when supplying new data, is not acceptable without clear reference to the process. One duplicate figure in a paper is one too many, if attribution to the original paper or grant is not noted. Oblique reference to a method in a previous publication in an attempt to hide the paper's intellectual precedents is still deceitful and a form of plagiarism.

Editors have an obligation to act if concerns are raised about improper attribution. If authors do not supply satisfactory explanations, their employers and funding agencies must be notified. It is the responsibility of institutions, who have a legal mandate, to initiate a formal investigation.

Timeliness can be difficult if institutes are reluctant to taint their reputations with negative findings, or if international boundaries are crossed. Editors should nudge investigations that drag, and draw attention to incidents where no satisfactory progress is made.

Where plagiarism is found, the author's previous publications must be examined. The evidence shows that an act of misconduct is usually part of a pattern of behaviour rather than an isolated incident, says Richard Smith, former editor of the *British Medical Journal*.

Journals should proceed promptly to correct the literature where discovery of misconduct necessitates it. Plagiarized text or figures should be clearly indicated as such within the original content. *Nature* will play its part where necessary, as will other *Nature* titles. One might hope that such public humiliation will act as a deterrent to those inclined to pass off another's work as their own. ■

"Editors are obliged to act if concerns are raised about improper attribution."

RESEARCH HIGHLIGHTS

Mouthing off

J. Virol. **79**, 13587–13593 (2005)

Some plant viruses are transmitted only by a particular insect. However, they retain the ability to evolve quickly into a form that can be passed on by a different species, suggests work by Stéphane Blanc, of the Montpellier INRA Centre in France, and his colleagues.

The cauliflower mosaic virus has a protein that binds it to the mouth parts of an aphid (pictured). It hitches a ride on the aphid's stylet, which pierces the plant when the insect feeds.

The researchers identify the region of the protein that binds to the stylet, and pinpoint a single amino-acid residue that determines the protein's affinity for different aphid species. The virus can swap vectors by changing just this one residue.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

V. STEGER/SPL

GEOPHYSICS

Intruders from the deep

Science **310**, 654–657 (2005)

Slow-spreading ridges on the ocean floor may incorporate older rocks from deep down into newly formed crust, say Joshua Schwartz, of the University of Wyoming in Laramie, and his colleagues.

The team performed radiometric dating on samples of the mineral zircon, collected from Atlantis Bank through submersible dives and dredging. This region is some 100 kilometres south of the ridge that separates the African and Antarctic plates. Around one-quarter of the samples turned out to be far older than is calculated by looking at the sea floor's magnetism — by up to 2.5 million years. The discrepancy can be explained if gabbroic rocks that crystallized

up to 18 kilometres beneath the ridge were lifted up and mixed into crust forming at the crest of the ridge.

DEVELOPMENT

Extra strong eggs

Ecol. Lett. **8**, 1105–1113 (2005)

Support for the idea that the speckles on eggshells provide structural support, rather than camouflage, comes from a study of great tits' eggs.

The great tit (*Parus major*), like many other passerine species, lays white eggs with red speckles (pictured). Researchers led by Andrew Gosler of the University of Oxford, UK, collected more than 100 such eggs. They found that the pigment appears in regions where the shell is thinner, providing additional reinforcement — the thinner the shell, the more dense the pigment. What's more, egg clutches produced in regions naturally high in calcium tend to have thicker shells and fewer spots. The pigment compounds, known as protoporphyrins, may act as lubricants between the calcite crystals that make up the eggshells, reducing brittleness.

GENETICS

Tracing a cell's family tree

PLoS Comput. Biol. **1**, e50 (2005)

It is theoretically possible to construct the family tree of each cell in the human body, according to researchers in Israel. The team, led by Ehud Shapiro at the Weizmann Institute of Science in Rehovot, says that naturally occurring mutations could be used

to trace the lineage of cells in a newborn baby back to the fertilized egg.

Shapiro and colleagues performed a mathematical study of the frequency of mutations in stretches of the genome known as microsatellites. They found that mutations occur often enough to track a cell's history back through 40 cell divisions.

Although genome sequencing technology is not yet advanced enough for the technique to be applied to whole people or even mice, the team hopes it will soon help to discover how cells in cancers grow and spread.

NEURODEGENERATION

Enzyme goes awry

Cell **123**, 277–289 (2005)

Mitochondria, the tiny bacterium-like structures that help to produce energy in our cells, have long been linked to a range of hereditary diseases. Now researchers in Germany and Italy have uncovered the role of one mitochondrial enzyme that is associated with progressive neurodegeneration.

A team led by Thomas Langer at the University of Cologne studied an enzyme called paraplegin, known to be faulty in patients with hereditary spastic paraplegia (HSP). Working in yeast and mouse cells, Langer and colleagues found that paraplegin controls the final step in the assembly of the mitochondrion's ribosomes, the molecular machines that make proteins.

In those suffering from HSP, the loss of axons, which carry nerve impulses, could be linked to faulty protein synthesis in their mitochondria.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

R. PACKWOOD/OXFORD SCIENTIFIC FILMS

CHEMISTRY

Laundromaths

Langmuir **21**, 10106–10111 (2005)

For most people, doing laundry is a chore; but for chemists it's an absorbing problem. Dinesh Shah and colleagues of the University of Florida in Gainesville are tackling the spin-drying cycle.

Detergents contain surfactants that lower the surface tension of water, reducing how strongly moisture is sucked between cloth fibres. Simple reasoning therefore suggests that you will get drier clothes if you increase the concentration of detergent in the water.

Unexpectedly, Shah and his team find that the water content of fabrics after a spin cycle actually rises when surfactant concentration is increased beyond a certain level. They attribute this behaviour to the adsorption of surfactant molecules on to the fibres, which reduces the concentration of the molecules in the water. This provides a clue to how detergents could be designed to make clothes dry faster.



process called reionization that makes the gas transparent to light. If so, this discovery will help to pin down the timing of the era of reionization — the end of what is known as the cosmic dark ages.

ASTRONOMY

When darkness lifted

Astrophys. J. **633**, L1–L4 (2005)

In the early Universe, the fog cleared quickly. That's the implication of an ancient galaxy detected by researchers using the Hubble and Spitzer Space Telescopes.

Bahram Mobasher of the Space Telescope Science Institute in Baltimore, Maryland, and his co-workers say that HUDF-JD2 is a massive galaxy that formed only a few hundred million years after the Big Bang, or as astronomers describe it, at a redshift of around 15. The light from the stars in this galaxy could have stripped electrons from the surrounding primordial gas, starting a

NEUROBIOLOGY

The nose knows

Nature Neurosci. doi:10.1038/nn1589 (2005)

It was thought that rodents detect pheromones — chemicals that influence their sexual and social behaviour — primarily through a specialized structure called the vomeronasal organ. But experiments in mice now show that the main olfactory system also plays a crucial role in picking up the signals that drive mating and aggression.

Male mice engineered to lack a gene that underlies odour detection in the main olfactory tissue showed no interest in female mice, report Nirao Shah and his colleagues at the University of California, San Francisco. The mutants also showed reduced aggression when confronted by other male mice.

DYSLEXIA

Reading the genome

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0508591102 (2005)

Researchers are homing in on the genetic causes of dyslexia, a complex reading disorder that affects 5–17% of the population and that seems to be highly heritable.

Previous studies have linked a region of chromosome 6 with a predisposition to dyslexia. Jeffrey Gruen, of the Yale Child Health Research Center in New Haven, Connecticut, and his colleagues examined this region in 153 families affected by reading disability. They all possessed the same deletion in the *DCDC2* gene. The exact function of the gene remains unknown, but when the researchers decreased levels of the *DCDC2* gene product in rats, some neurons in the rodents' brains failed to develop properly.

CELL BIOLOGY

Building site

Science doi:10.1126/science.1119969 (2005)

Just as cells must replicate their chromosomes when they divide, so too must they duplicate bodies in the cytoplasm. A chance observation now provides insight into how the cytoplasm's Golgi apparatus, which processes and distributes proteins, is copied.

Graham Warren and colleagues from Yale University in New Haven, Connecticut, used antibodies to label centrin proteins in the single-celled parasite *Trypanosoma brucei*. This revealed a structure near the Golgi that contained centrin2. This protein had previously been seen only in the centrosome — which coordinates the growth of some of the cell's skeletal elements. The structure may provide the site at which the new Golgi is assembled.

JOURNAL CLUB

Anthony Ryan
University of Sheffield, UK

Britain's ICI professor of physical chemistry gives praise to his competitors.

Don't you just hate it when you read a paper and wonder "Why didn't I think of that?"

This was my reaction on reading Michael Massa and Kari Dalnoki-Veress's work in *Physical Review Letters* (**92**, 255509; 2004). They present a technique to study

nucleation — the way that crystallization starts. It is elegantly simple compared with the brute-force method that my team has been using.

Crystallization defines the properties, both mechanical and aesthetic, of most plastics we use every day. And as polymers find more and more uses in nanoscale electronic devices and flexible displays, controlling polymer crystallization will become ever more important.

In small volumes of polymer, crystallization will start

spontaneously rather than at the site of a defect or impurity, as happens in large volumes. My team has been investigating how crystallization proceeds in finely divided matter using block copolymers that create nanometre-sized spheres of polymer.

But our progress in studying spontaneous nucleation has been laborious, because each time we wanted to change the volume of the spheres we had to synthesize new block copolymer molecules.

Massa and Dalnoki-Veress, both of McMaster University in

Ontario, Canada, instead used the process of dewetting — where a thin liquid film breaks up into droplets — to create an array of thousands of microdroplets with a range of sizes.

The researchers go on to show that the theory of spontaneous nucleation is valid right down even at the molecular scale.

Although disappointed that we didn't think to combine dewetting and crystallization (we study both), I am consoled by the fact that the authors quote some of our work as their inspiration. Thanks!

NEWS

Wartime tactic doubles power of scarce bird-flu drug

Doctors think they have hit on a way to effectively double supplies of a drug that fights bird flu. Administering Tamiflu alongside a second drug that stops it being excreted in urine means that only half doses of the treatment would be needed.

Tamiflu (oseltamivir phosphate) is the main anti-flu medicine recommended by the World Health Organization (WHO). The WHO suggests that, in anticipation of a flu pandemic, countries should stockpile enough for at least a quarter of their population. But although Swiss drugmaker Roche, the sole supplier, has quadrupled its production capacity over the past two years, the current supply is thought to cover just 2% of the world population.

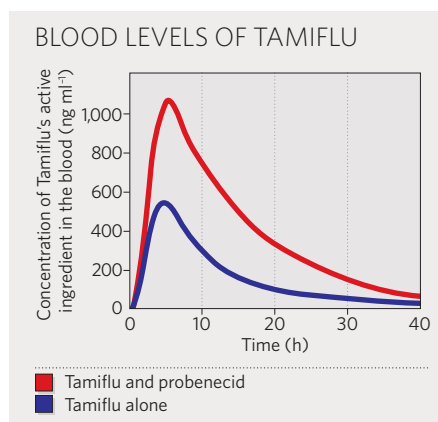
Last week, Joe Howton, medical director at the Adventist Medical Center in Portland, Oregon, suggested a way to double supplies, after browsing basic safety data from Roche for a talk on avian flu.

The technique was invented during the Second World War to extend precious penicillin supplies. Scientists found that a simple benzoic acid derivative called probenecid stops many drugs, including antibiotics, being removed from the blood by the kidneys. Probenecid is readily available and is still widely used alongside antibiotics to treat gonorrhoea and syphilis, and in emergency rooms, where doctors need their patients to have high, sustained levels of antibiotics in their blood.

Howton noticed from Roche's data that Tamiflu, like penicillin, is actively secreted by the kidneys, and that the process is inhibited by probenecid. Giving the flu drug together with probenecid doubles the time that Tamiflu's active ingredient stays in the blood, doubles its maximum blood concentration, and multiplies 2.5-fold the patient's total exposure to the drug (see graph, and G. Hill *et al. Drug Metab. Dispos.* 30, 13–19; 2002).

In other words, you could get away with using half as much Tamiflu to get the same therapeutic effect. "It dawned on me that the data potentially represented a tremendous therapeutic benefit," Howton told *Nature*.

Given that Roche published the probenecid data in 2002, has it considered this option? "It doesn't seem so," says Martina Rupp, a spokeswoman at Roche's headquarters in Basel. "It is an interesting idea, but we can't really say anything," she adds, claiming that



there are insufficient data. The WHO and the US Food and Drug Administration declined to comment when *Nature* asked them about the idea.

Studies are being proposed that will look at safety issues relating to probenecid and Tamiflu, although doctors argue that there are already enough data for the drug combination to be used, even without specific approval from regulatory agencies. Gratian Woodson of the Atlanta Research Center in Decatur, Georgia, has prescribed probenecid for more than 25 years and says he prescribes drugs for such off-label purposes every day. "This is a perfectly acceptable and established practice," he says.

Peter Zed, a specialist in emergency medicine at Vancouver General Hospital in Canada, agrees. He has published studies of the safety

of probenecid and antibiotic combinations. "There would be nothing unique about using probenecid with Tamiflu," he says.

Michael Osterholm, director of the US Center for Infectious Disease Research and Policy in Minneapolis, Minnesota, cautions that probenecid alone will not be sufficient to avert a flu pandemic. He points out that the most optimistic estimate of Tamiflu production capacity in the next five years gives enough to treat just 7% of the global population.

Coping with a pandemic will require "launching a worldwide Manhattan-like project for drug production, packaging and distribution today," Osterholm says. "It's not just about having a magic bullet; it's whether you can make it and find enough guns from which to shoot it." Still, doubling the doses available could be crucial for treating people quickly after an outbreak, and Osterholm says the idea definitely merits investigation.

"It's not just about having a magic bullet; it's whether you can find enough guns from which to shoot it."

"This is wonderful," agrees David Fedson, formerly a medical director of the vaccine company Aventis Pasteur, based in Lyons, France. "It is extremely important for global public health because it implies that the stockpiles now being

ordered by more than 40 countries could be extended, perhaps in dramatic fashion." He suggests that capsules containing both Tamiflu and probenecid should be developed.

Like many scientists, Fedson is stumped by the apparent lack of interest from Roche, and the relevant authorities. "It's stupefying," he says. ■

Declan Butler

Drug firms donate compounds for anti-HIV gel

Motivated by positive results reported in this week's *Nature*, two drug companies have given away rights to two key compounds, so that they can be developed into gels that protect against HIV.

Such a gel could help many women to protect themselves, as they often find it difficult to get partners to use condoms —

particularly in the developing world, where men may disapprove of the practice. Experts say that a microbicide applied to the vagina before sex could save 2.5 million lives in just three years.

But progress to develop such gels has been slow. Only one microbicide trial has been completed in humans, with

disastrous results — the women became more susceptible to HIV because the gel, essentially a detergent that destroys the virus, damaged their vaginal tissue. Five other microbicides are in clinical trials in Africa after proving moderately successful in monkeys, but critics point out that the virus used in those animal tests

SOURCE: G. HILL ET AL. *DRUG METAB. DISPOS.* 30, 13–19 (2002)



SMALL SPARKS PACK A BIG PUNCH
Sparks made in the lab offer clues to how lightning works
www.nature.com/news

Universities scramble to assess scope of falsified results

Biologists are rushing to quantify the fallout from a case of scientific misconduct unmasked last week. The Massachusetts Institute of Technology (MIT) confirmed on 27 October that it had fired immunologist Luk Van Parijs for fabricating and falsifying data.

The institute said that Van Parijs, an associate professor of biology, had acknowledged to its officials that he altered data in one published article, in unpublished manuscripts and in grant applications. The California Institute of Technology (Caltech) and Harvard University have both now opened inquiries into some of Van Parijs's other published work.

Authorities at all three universities say they have found no evidence that anyone else was involved in the misconduct. Van Parijs previously worked in the labs of Caltech's president, David Baltimore, and physician Abul Abbas, head of pathology at the University of California, San Francisco.

Van Parijs, a 35-year-old native of Belgium who lives in Falmouth, Massachusetts, did not respond to interview requests. His primary studies involved using short pieces of RNA to silence genes that have gone awry in autoimmune diseases. Early indications suggest that his misconduct will not affect his field as dramatically as semiconductor research was affected by Bell Laboratories physicist Jan Hendrik Schön¹. Colleagues spent years following Schön's line of research until in 2002 it was discovered that he had falsified some of his data.

Van Parijs co-authored a heavily cited paper for *Nature Genetics* in 2003 describing how to use a stripped-down virus, called a lentivirus, as a delivery system for genes that can silence other genes². The paper has not been called into question, and other researchers have shown that the approach works, at least in animal models and cell cultures. Clinical trials in human patients are now being planned.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sacked: Luk Van Parijs, seen as a rising star in biology, was found to have falsified data.

Because of such work, Van Parijs was considered a rising star. "He had incredible potential," says Abbas, who worked with him in 1997 and 1998 at Brigham and Women's Hospital in Boston. "He was a superstar in the making."

But in August 2004, MIT put Van Parijs on administrative leave after some of his lab colleagues made allegations of misconduct to the institute's authorities. After the lab was closed, his students and postdocs scrambled to get jobs elsewhere, and MIT began the inquiry that culminated in his sacking.

Abbas says that Van Parijs has e-mailed him

and denied falsifying data in more than one paper. "He apologized for the disappointment he has caused," Abbas says.

MIT officials say that they will submit the results of their investigation to the Office of Research Integrity, the federal agency that monitors research conduct for the US National Institutes of Health. It has been at least a decade since the institute has uncovered a case of misconduct, the officials say.

MIT has not publicly identified the paper that contains the falsified results. But in May, *Current Opinion in Molecular Therapeutics* published a correction to a 2004 review^{3,4} on which Van Parijs was the lead author. The note said that unpublished experiments cited in the paper — involving genetically controlling tumour growth in mice — could not be documented.

Investigators are now probing several of Van Parijs's older publications. Caltech is looking at two articles published in *Immunity*, including one co-authored by Baltimore^{5,6}. And Harvard is looking into a 1997 paper in the *Journal of Experimental Medicine*⁷.

The *Immunity* work dealt with a cell signalling pathway that governs the processes by which cells in the immune system live and die. An expert in the field, who asked not to be named, said that Van Parijs's experiments had never directly been replicated. "We really would like to know if the work is reproducible, so the field can move forward," the immunologist said, "or whether we have to do an about-face." ■

Rex Dalton

Additional reporting by Erika Check

1. Brumfiel, G. *Nature* **419**, 419–421 (2002).
2. Robinson, D. A. et al. *Nature Genet.* **33**, 401–406 (2003).
3. Nencioni, A. et al. *Curr. Opin. Mol. Ther.* **6**, 136–140 (2004).
4. *Curr. Opin. Mol. Ther.* **7**, 282 (2005).
5. Van Parijs, L., Peterson, D. A. & Abbas, A. K. *Immunity* **8**, 265–274 (1998).
6. Van Parijs, L. et al. *Immunity* **11**, 281–288 (1999).
7. Van Parijs, L. et al. *J. Exp. Med.* **186**, 1119–1128 (1997).

infects cells in a different way from the one that causes AIDS.

John Moore from Cornell University in New York and his colleagues tried a different approach (see page 99). They combined three compounds that each uses a different mechanism to block the virus's entry into cells. Merck's compound CMPD167 competes with the virus for cell

receptors inside the vagina. Bristol-Myers Squibb's BMS-378806 interacts with the virus itself, stopping it binding to cells. And a peptide developed by Moore's team inhibits the process used by the virus to enter a cell.

When the researchers tested combinations of the compounds in macaques, they found that they offered at least partial protection

against a virus closely resembling HIV. But three animals that received the three compounds together were all protected against infection. These results were enough to persuade the drug firms to give away rights to the compounds, says Moore. "This is the first time there has been a joint announcement like this," adds Mark Mitchnick, chief scientific officer of the International Partnership for

Microbicides, the non-profit group that will develop the gel.

Partners including the Bill and Melinda Gates Foundation and the US National Institutes of Health are helping to fund a clinical trial, set to start in 2007. This is estimated to cost between US\$150 million and \$200 million and will involve about 10,000 women in Africa. ■
Narelle Towie

ON THE RECORD

“Perhaps what we should do is give each person a gun, and when we see a migrating bird, we can just shoot it down.”

Hong Kong lawmaker Tommy Cheung offers an unusual proposal for controlling the spread of avian flu.

“We issued a general recommendation for poultry producers to prevent the spread of the virus with the sauna.”

Sirpa Kiviruusu from Finland's agricultural ministry backs steam rooms as a disinfection method for controlling the bird flu virus.

Sources: AP, Agence France-Presse

SCORECARD

**Deer**

In Missouri, the tables have turned and this year deer are doing the shooting. Researchers have so far collected 200 hours of film from cameras mounted on the animals to give a deer's-eye view of the world.

**Wasps**

Researchers in Georgia are creating a buzz with a prototype detector for explosives. The device uses wasps trained to react to certain chemicals in the air.

**Pluto**

As the debate about Pluto's planetary status rages on, two extra moons spotted orbiting the body may help boost its image.

DATAPOINT

Some countries benefit from a 'brain drain' of qualified doctors from elsewhere. In Britain, for instance, a recent study finds that 28% of physicians come from abroad. The top four suppliers are:

India	10.9%
Ireland	2.1%
Pakistan	1.9%
South Africa	1.4%

Source: New England Journal of Medicine

Protists push animals aside in rule revamp

Organisms whose cells have a nucleus — eukaryotes — have traditionally been separated into four 'kingdoms'; now they have been reorganized into six. The authors of the revision hope that it will bring peace to a long-divided discipline, and raise awareness of the diversity of single-celled organisms.

Textbooks generally divide eukaryotes into plants, animals, fungi and protists. The protist kingdom mostly consists of single-celled organisms such as amoebae. Bacteria make up a fifth kingdom.

Since the late 1970s, data from electron microscopy and DNA sequences have indicated that the traditional groupings do not make sense. Protist classification was particularly troubled. But the evidence was not clear enough for a consensus on a new regime, and fierce disagreements became common.

"I've seen people throw things at each other," says Sina Adl, a soil-organism specialist at Dalhousie University in Halifax, Canada, who coordinated the group of 28 protist experts that produced the new classification. It was commissioned by the International Society of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Trypanosomes (purple), which cause sleeping sickness, were one of the hardest protists to classify.

Protistologists and is published in *The Journal of Eukaryotic Microbiology* (S. M. Adl *et al.* *J. Eukaryot. Microbiol.* **52**, 399–451; 2005).

The experts have given protists a mighty four kingdoms out of six. Animals do not even get their own group — fungi and animals have

Turkish rectors rally in support of university head thrown in jail

The struggle between Turkey's Islamic and academic powerhouses for control of the country's universities is becoming increasingly acrimonious.

Late last month, nearly all of the country's 77 university rectors travelled to Van, close to Turkey's eastern border with Iran, to support Yücel Askin, a university rector who is held in jail there. They say that the charges brought against him are trumped up, and are demanding his release.

Askin, who has been rector of the 100th Year University in Van since 1999, has been charged with corruption relating to the purchase of medical equipment for the university hospital.

According to Turkish law, all problems arising in the country's universities should

be handled first by the Council of Higher Education. But the public prosecutor in Van ordered Askin's arrest on 14 October without asking the council, and insisted that he is jailed until a date for the court hearing is decided.

This is the fifth case that the public prosecutor has tried to bring against Askin. The others involved accusations of illegal handling of historical documents or objects snatched from his home during police raids in the past year. All items were found to be legally registered and the charges dropped.

Askin is unpopular with religious groups, say the rectors, who argue that he has been targeted because he resisted pressure to appoint underqualified candidates,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

been merged into Opisthokonta, and plants are called Archaeplastida.

As for the protists, amoebae and slime moulds form Amoebozoa, and various single-celled organisms are now Rhizaria. The remaining two groups — Chromalveolata and Excavata — are the most contentious. “There have almost been fist fights over the mention of these groups,” says Adl. “It has taken a lot of

diplomacy to get people to sit down and talk.”

The tensions arise from conflicting DNA evidence. Some genes are different enough to suggest that the groups should be separate kingdoms, others are not. Adl’s team decided that enough evidence had accumulated to declare them distinct. “A lot of people will be upset,” he admits. “But it needed doing — over the past 20 years the classification we had has fallen apart.”

One dissenter is evolutionary biologist Blair Hedges, based at Pennsylvania State University in University Park. He says genomics work by himself and others suggests that the Excavata should not be a separate group. “People have ignored the evidence and gone on gut feeling,” he told *Nature*.

The two contentious groups include parasites that cause diseases such as malaria and sleeping sickness; Adl hopes his classification will aid drug development. “Placing these organisms in the wrong group is in part responsible for the fact that we do not have specific drugs for these diseases, because of wrong assumptions about their biochemistry,” he says.

But Michael Gaunt, who studies the Chagas’ disease parasite at the London School of Hygiene and Tropical Medicine, is less convinced that the change will have dramatic practical effects. “It is important to understand the relationships between these organisms if we are to tackle these diseases,” he says. “But the lack of effective drugs is largely due to poor industrial interest.” ■

Tom Simonite

EYE OF SCIENCE/SPL

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

University rectors from across Turkey head for Van to protest over the treatment of Yücel Askin.

supported by powerful Islamic figures, into key academic positions.

Turkey’s higher-education law is supposed to forbid political or religious interference in academia. But tensions

between Islamic and secular groups have come to a head since 2003, when the country’s centre-left-led coalition was replaced by the Justice and Development Party, which is mildly Islamic (see page 1).

The rectors flew to Van on 23 October to register their concern and investigate the issue. During the visit their 20-car convoy was attacked by demonstrators described by the rectors as fundamentalist Islamics.

Erdogan Tezic, president of the higher-education council, told national newspapers that Askin wanted “to maintain the secular quality of the republic”, and that “defending Mr Askin is synonymous with defending the Republic”. The Turkish Academy of Sciences has also released a statement in support of Askin.

Justice minister Cemil Cicek has told newspapers that the incident should not be turned into an ideological cause, and that the courts will decide the outcome. In the meantime, the case remains at an impasse. ■

Alison Abbott

U. KOZAN/DHA

Floods fail to save canyon beaches

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

D. MUENCH/CORBIS

The Colorado River has been deprived of silt by a dam — with fatal consequences for local fauna.

PHOENIX, ARIZONA

River habitats along the Grand Canyon, which have been eroding because of a dam upstream, may never return to their natural state, admit Earth scientists. In a report released last week, they conclude that floodwater experimentally released into the river did not carry enough sediment to rebuild the area's eroded beaches.

The report reviews more than ten years of work, and was released at the Colorado River Ecosystem Science Symposium, held in Tempe, Arizona, from 25 to 27 October. In the study, the researchers say that a flood released in 2004 created some sandbars in the upper reaches of the river, but admit the effect was short-lived.

The Colorado River, which runs along the Grand Canyon in Arizona, used to be brown with sediment. It was flanked by extensive beaches and sandbars that provided habitat for species found nowhere else. The humpback chub (*Gila cypha*) is one such species — a fish with degenerate eyes and a prominent hump that stabilizes it in the river's swirling waters.

However, in 1963 the Glen Canyon Dam, which is now the major power source in the

southwestern United States, was completed. As a result, the river downstream of the dam runs slow and clear, the beaches and sandbars have eroded dramatically, and the humpback chub is critically endangered.

In 1996, after years of debate, scientists at the US Geological Survey agreed on a plan to restore the ecosystem. They hoped that releasing a massive flood through the dam would shift sediment trapped at the bottom of the river and rebuild the beaches. Later that year, they pumped an extra 1,290 cubic metres of water per second down the canyon for a week, at an estimated cost to power companies of \$2.5 million. It was a failure — there just was not enough sediment in the river to redistribute (see *Nature* 420, 356–358; 2002).

In 2004 they tried again, timing the flood to coincide with an annual release of sediment from the Paria River, just downstream of the Glen Canyon Dam. But despite high hopes, this failed too. The results “directly contradict” rosy predictions made in 1996, says the scientists' report, which concludes that the only way to rebuild the sandbars permanently is with a continuous source of new sediment.

“Full restoration of the river is impossible,” says John Schmidt, a geomorphologist at Utah State University in Logan, who helped to plan the 1996 and 2004 floods. “It is a much more difficult challenge than anticipated.” He says it's time for Americans to realize that they cannot have a huge power-generating dam and a natural river ecosystem.

But others continue to look for solutions. The US Bureau of Reclamation, which operates the dam, is thinking about piping sediments into the Colorado River from a lake that has been created upstream of the installation, Lake Powell. A sediment transport system would be a momentous engineering endeavour, estimated to cost at least \$100 million to build and a further \$10 million a year to operate. Even if the project is funded, it may take a decade to complete.

In the meantime, efforts to save the humpback chub, including culling the invading trout that feed on it, are not working. Scientists at last week's symposium presented results showing that more than half of the 10,000 chub that thrived a decade ago are now gone. ■

Rex Dalton



THE NATURE PODCAST
Listen to the stories behind the hottest research in our weekly show, available free at www.nature.com/nature/podcast

Expert witness: the scientists who testified against intelligent design

In 2004, a school board in Dover, Pennsylvania, announced that biology teachers would have to read out a statement about evolution in their lessons. The wording refers to “gaps” in Darwin’s theory and promotes intelligent design, the concept that a creative intelligence guides evolution. A group of parents filed suits against the board and called on two scientists to testify in support of darwinian theory.

Geoff Brumfiel sat down with Kenneth Miller, a cell biologist, and palaeontologist Kevin Padian to find out about their day in court.

Why were you chosen to be expert witnesses?

Miller: I wrote the textbook that they use in Dover. When this situation arose, I sent an e-mail to several of the Dover biology teachers saying I’d be happy help out. A month or two later the lead attorney for the plaintiffs called me up and asked me if I’d want to be an expert witness.

Padian: I’m president of the National Center for Science Education, a non-profit organization based in Oakland, California, that is particularly interested in the creationism–evolution debate.

Have you ever been called to do this sort of thing before?

Miller: I’ve appeared in a courtroom one other time, during a trial in Atlanta, over warning stickers that some of the school districts in Georgia had tried to put on my textbook. But I was not an expert witness, I merely testified to factual questions about how the book was written.

Padian: This is my first time.

Why did you feel it was important to testify?

Padian: It’s an opportunity when it really counts. One person can’t be everywhere around the country talking to every school board and every parent group. But this is a case where, ultimately, these decisions are going to clarify things in a formal setting.

Miller: It is the right thing to do. The battle in Dover is just one example of local battles for scientific education all over the country. If people in the scientific community turn their backs on people in the front lines, then

What they said in court

Kenneth Miller
Cell biologist at Brown University in Providence, Rhode Island

“If you invoke a non-natural cause, a spirit force or something like that in your research and I decide to test it, I have no way to test it. I can’t order that from a biological supply house, I can’t grow it in my laboratory. And that means that your explanations in that respect, even if they were correct, were not something I could test or replicate, and therefore they really wouldn’t be part of science.”

ultimately the cause of science in public education is doomed.

How long did it take to prepare for the trial?

Miller: Through the spring and summer I devoted 100–150 hours to preparing for this case.

What did you have to do to get ready?

Padian: I had to prepare an expert report, which basically says what you will talk about at the trial. It’s kind of like a review paper, except it has to be comprehensible to the judges and lawyers. Then I had to undergo a deposition, when the other side interviews you to find out what you’re going to say.

You were interviewed by the lawyers for the school board. What was that like?

Padian: It wasn’t really a hostile situation. The lawyer for the other side, a guy whom I like and respect very much, was really just trying to get his head around the science of evolution. **Miller:** They tried to make it friendly, but it went on for nine-and-a-half hours, and it was pretty gruelling in the end.

What was the day of the trial like?

Miller: Like most scientists, I’m a stranger to

Kevin Padian
Palaeontologist at the University of California, Berkeley, California

“We’ll be the first people to admit that science doesn’t know everything and can’t know everything. But on the other hand, we would like a fair and accurate representation of what we do know.”
“It worries me that students would be told that they have to make a conclusion in advance of all the evidence that you can’t get from A to B, essentially, by natural means.”

the courtroom, and I found the decorum, the formality and the attention to procedure really striking. Arguments were heard and heard fairly. I was impressed with that.

Padian: I didn’t feel nervous. The judge was great. He is very smart, he is very attentive and he is running a tight ship, but it’s not stiff. The defence lawyers are good people, and our counsel is just fabulous. The testimony took most of the day, and I was glad to relax after a process that had taken months of preparation.

What can other scientists learn from your experience?

Miller: Science in a modern society is ultimately dependent on the public for support and acceptance. Scientists everywhere and in every field must be willing to make the case for science to the general public.

Padian: If you want to explain something to people, it has to be put in terms of the issues that they find important. Politicians, judges and the media are not impressed by someone thundering in and claiming that they have all the right answers.

As *Nature* went to press, arguments in the case were expected to conclude on 4 November, although a decision is not likely before December.

Gene study raises fears for three-parent babies

SALT LAKE CITY, UTAH

A dramatic controversy has erupted over assisted-reproduction techniques that mix mitochondria from different women.

Mitochondria are the energy-generating structures in cells. Adding mitochondria from the egg of a second woman during *in vitro* fertilization is thought to boost the development of some eggs. It could also correct inherited diseases in which the mitochondria are defective.

But the approach is surrounded by ethical concerns, because the resulting babies have DNA from two mothers as well as the father. No teams are using the technique to treat women now, although such babies have been born in the past, and UK researchers are using it to create human embryos (see *Nature* 437, 305; 2005).

Now a mitochondrial researcher has reported work that he says raises health worries too. "Our results suggest we shouldn't mix mitochondrial DNA from different women without thinking through the consequences,"

says Douglas Wallace, a biologist at the University of California, Irvine. Wallace described his research, on genetically engineered mice, on 26 October at a meeting of the American Society of Human Genetics in Salt Lake City, Utah.

Assisted-reproduction specialists say Wallace's conclusions are too rash. "To compare these artificial animals to assisted reproduction in humans is simply not correct," says Jacques Cohen, research director of the company Reprogenetics in West Orange, New Jersey.

Wallace's concerns come from a decade of research. Mitochondria carry genes that are passed unchanged from mother to child. But when Wallace's group created mice with mitochondria from two mothers, the mitochondria swapped genetic material in the offspring. Wallace says these mice had fewer pups — implying an effect on fertility and possibly on health.

Wallace is worried that assisted reproduction

may subvert a natural system that protects mitochondria from changes that could perturb their function. "Mitochondrial DNA programs are highly adapted and integrated. It's extremely bad planning to put those different programs in the same cell," Wallace says.

Other scientists are sceptical that DNA from different mitochondria would mix. Brendan Battersby of McGill University in Montreal, Canada, has also made mice with mixed mito-

chondria, and says he has never seen it happen: "I just don't believe it."

James Grifo of the New York University Fertility Center, who used mitochondrial transfer to create several babies in the late 1990s, cautions that it is too early to draw conclusions from Wallace's results. "I'm not sure how they can make such broad statements about this being dangerous," he says. ■

Erika Check

Physicists denounce aggressive nuclear policy

More than 700 physicists from around the world have signed a petition opposing a US policy that would permit the use of nuclear weapons against non-nuclear nations.

Spawned during a lunchtime talk at the University of California, San Diego (UCSD), the petition is being submitted to US government leaders. Eight Nobel laureates have signed the petition, which was started by UCSD physicists Kim Griest and Jorge Hirsch.

The administration of President George W. Bush has said that, if provoked, it would consider using nuclear bombs on a country without such weapons.

"Physicists were responsible for these weapons," says Hirsch, a native of Argentina. "We need to speak out more."

The petitioners hope to win the support of the American Physical Society and the International Atomic Energy Agency at board meetings later this month.

► <http://physics.ucsd.edu/petition>

Fire engulfs research centre at Southampton University

Fire has destroyed a leading UK optoelectronics research centre.

Firemen took nearly 10 hours to bring a blaze that started on 30 October under control at the University of Southampton's Optoelectronics Research Centre. Initial reports suggested little would be salvaged from the building, which included a microfabrication facility. The fire also destroyed laboratories and offices of the School of Electronics and Computer Science.

University officials were surveying the damage as *Nature* went to press and said that further announcements would be made this week. No injuries had been reported and the cause of the fire is not thought to be suspicious.

Kansas raises hackles with rewrite of science guides

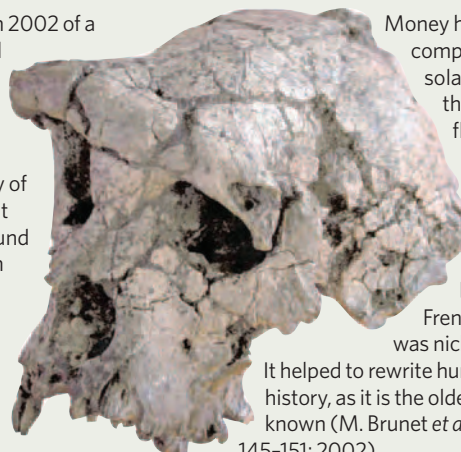
The US science establishment is washing its hands of education standards in Kansas that portray evolution as "controversial" and remove wording that defines science as "systematically seeking natural explanations for what we observe in the world around us".

The Kansas standards are based on copyrighted guidelines that were originally prepared by the National Academy of Sciences (NAS), based in Washington DC, and the National Science Teachers Association (NSTA) in Arlington, Virginia.

Chad gives fossil hunters a head start

Inspired by the discovery in 2002 of a 7-million-year-old hominid skull within its borders, Chad last month opened its first palaeontology department.

Housed at the University of Ndjamena, the department will serve as a training ground for scientists from western and central Africa, says its head, Hassane Taisso Mackaye. The newly refurbished building will also be home to a collection of vertebrate skeletons from the Sahel.



Money has been spent on computers, as well as solar panels to keep the electricity flowing, as Ndjamena suffers from regular power cuts. The hominid skull discovered by Chadian and French researchers was nicknamed Toumaï.

It helped to rewrite human evolutionary history, as it is the oldest hominid known (M. Brunet *et al. Nature* **418**, 145–151; 2002).

Nearly every state uses some version of the guidelines in their science standards. But the Kansas Board of Education edited the text to conform to the anti-evolution sympathies of six of the ten board members.

After looking over the standards, the academy and the teachers' association last week refused to grant permission to use their source materials, saying the wording compromised the original documents.

The board will vote on the edited standards on 8 November. They may be approved, with the understanding that passages will be rewritten to avoid infringing the NAS and NSTA copyrights, says board secretary Penny Plamann.

NASA gets go ahead to buy Russian trips to space

NASA came closer to solving its transportation woes last week with the passage of legislation that will let US astronauts continue to ride on Russian spacecraft.

The House of Representatives passed amendments to the Iran Nonproliferation

Act of 2000, which had prevented the United States purchasing seats on the Soyuz vehicle that currently serves as the lifeboat on the International Space Station (see *Nature* 436, 11; 2005). Russia had been providing seats as part of the space-station partnership; NASA was supposed to take over with its own rescue service when that agreement expired, but a planned US vehicle was cancelled in 2001.

The amendments, which are expected to pass the Senate and be signed into law by President Bush, allow NASA to buy the Russian services it needs to keep operating the space station until 2012. By that time, the United States is expected to have a new Crew Exploration Vehicle as a replacement for the space shuttle. NASA may also be booking commercial flights to orbit by 2012; the agency plans in December to solicit industry ideas for providing transportation to the station.

Stock ownership of FDA ex chief comes to light

A possible reason for the sudden departure of the US Food and Drug Administration's (FDA's) chief is starting to emerge.

Lester Crawford, who resigned abruptly in September without explanation (see *Nature* 437, 606; 2005), held stock in several companies regulated by the FDA while he was its top official, the government disclosed last week. He and his wife had stock in companies including Embrex, an agricultural biotechnology firm; Kimberly-Clark and Teleflex, which make medical products and devices; and Pepsico and Sysco, which make and distribute foods regulated by the FDA.

Crawford has not publicly described the full reasons for his departure, but some US newspapers have reported that the stock ownership may have played a role.



Ticket to ride: US astronauts will reach the space station by buying seats on Russian spacecraft.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Personal effects

Living things from bacteria to humans change their environment, but the consequences for evolution and ecology are only now being understood, or so the 'niche constructivists' claim. **Dan Jones** investigates.

In the Negev Desert of Israel, small organisms can have a big impact. Take the cyanobacteria that live in the soil. Some species secrete sugary substances that form a crust of sand and soil, protecting the bacterial colonies from the effects of erosion. When the rains come, the crusty patches divert water into pools in which wind-borne seeds can germinate. These plants in turn make the soil more hospitable for other plants. Thanks in part to these bacteria, patches of vegetation can be found where they might not otherwise exist. The action of the bacteria, together with local climate change, could lead to the greening of large parts of the desert.

The Negev cyanobacteria, and organisms like them, are also having an impact on evolutionary biologists these days. Examples of creatures altering their environment abound — from beavers that dam streams and earthworms that enrich the soil to humans who irrigate deserts. But too little attention has been given to the consequences of this, say advocates of niche construction. This emerging view in biology stresses that organisms not only adapt to their environments, but also in part create them. The knock-on effects of this interplay between organism and environment, say niche constructivists, have generally been neglected in evolutionary models. Despite pointed criticism from some prominent biologists, niche construction has been winning converts.

"What we're saying is not only novel, but also slightly disturbing," says Kevin Laland, an evolutionary biologist at the University of St Andrews in Fife, UK, and one of the authors of the idea¹. "If we're right, it requires rethinking evolution."

The conventional view of evolution sees natural selection as shaping organisms to fit their environment. Niche construction, by contrast, accords the organism a much stronger role in generating a fit by recognizing the innumerable ways in which living things alter their world to suit their needs. From this perspective, the road between organism and environment is very much a two-way street.

The intellectual stirrings of niche construction date back to the early 1980s, when Har-

vard University geneticist Richard Lewontin depicted evolution as a continual feedback loop, in which organisms both adapt to their environments and alter them in ways that generate new selective pressures. Although Lewontin's equations provided a broad perspective rather than a detailed model, he helped to kick-start the niche-constructivist approach, says Laland. "He really put the idea on the map."

Sons of soil

But it has taken years for biologists to begin to incorporate niche construction into more detailed models of evolution and ecology, in part because organism-environment interactions can be so complex. Earthworms, for instance, not only aerate the soil by tunnelling, as any gardener knows, but they also alter its chemical composition by removing calcium carbonate, adding their mucus and excrement, and pulling leaves down into the soil to decay. All of this produces a more favourable environment for worms to live in. Yet classical evolutionary models have typically failed to consider how this transformation alters the selective pressures on the worms and other soil inhabitants, say niche-construction advocates.

Back in the Negev Desert there are further examples of dramatic niche construction. At least three species of snail feed on lichens that live just below the surface of porous rocks. To

"What we're saying is not only novel, but also slightly disturbing. If we're right, it requires rethinking evolution." — Kevin Laland

vard University geneticist Richard Lewontin turned to differential equations — stock in trade for population biologists — to look at evolution from two different perspectives². He created one set of equations to describe the conventional view of evolution, the one-way-street version. A second set of equations, which he felt better described real evolutionary

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

From dissolving desert rocks to building dams, all organisms mould their environment to a certain extent.

get at the lichens, the snails have to literally eat through the rock, which they then excrete, creating soil around the rock in the process. This might sound insignificant, but it has been calculated that the combined action of these snails could generate enough soil to affect the whole desert ecosystem³. By transferring nitrogen in rocks to the soil, where plants use it for growth, the snails contribute substantially to sustaining local biodiversity.

Bigger picture

In extreme cases, niche-constructing activities can affect the whole world. The classic example from early evolutionary history is that of oxygen-producing cyanobacteria, which helped to set the stage for the evolution of animals and plants. Today, niche construction by human threatens to affect practically all life, as we pump large amounts of carbon dioxide into the atmosphere.

Critics are quick to point out that such cases have been well known to biologists for some time. "Darwin realized that organisms can change their environments in ways that affect their own evolution," says Laurent Keller, an evolutionary biologist at the University of Lausanne in Switzerland. "There are already many cases of niche construction by animals and especially humans," he says.

But advocates of niche construction counter that previous attempts to include these effects in evolutionary models have not gone nearly far enough. "People hadn't thought through the consequences of these effects, either for evolution or ecology," says John Odling-Smee, a biological anthropologist at the University of Oxford, UK.

To encourage people to consider the issue, Odling-Smee and Laland have taken a two-pronged approach. First, they have catalogued

hundreds of examples, involving thousands of species such as the Negev Desert organisms, to drive home the point that niche construction is a widespread phenomenon. In addition, they have developed mathematical models that capture the bidirectional nature of the niche-constructivist view, to show how these processes can actually be modelled.

Traditional ecological models typically distinguish between living things and their physical environment, but it is hard to model both elements at the same time. To find a way around this, Laland and Odling-Smee teamed up with Marcus Feldman, an evolutionary biologist and mathematical modeller at Stanford University in California. They found that they could look at niche construction by treating both living and non-living components of a niche as environmental factors that are both affected by, and feed back to, all the organisms in the ecosystem. They presented their results in a 2003 book¹, whose purpose, they say, was in part to convince other scientists to take niche construction into account in their research.

Perhaps the most direct way an organism

"Even Darwin realized that organisms can change their environments in ways that affect their evolution." — Laurent Keller

can alter the challenges it must face is by selecting where it lives, says Robert Holt, an ecologist at the University of Florida in Gainesville. Such habitat selection defines the future context for the evolution of the new residents and their progeny. By choosing to live in places to which they are already adapted, organisms can short-circuit the

selective forces that ordinarily lead to evolutionary change. In this way, habitat selection can lead to niche conservatism, which is the tendency not to adapt to new environments, and may explain the evolutionary stasis often seen in the fossil record.

Organisms can also shape their interaction with the world in more subtle ways. Developmental biologists know, for instance, that the mature form of many organisms varies depending on the environment in which they grow up. This is known as phenotypic plasticity. Although some creatures, such as beavers and cyanobacteria, alter their environment directly, others niche construct by modifying themselves, says Sonia Sultan, a botanist at Wesleyan University in Middletown, Connecticut. Sultan defines a niche according to the way an organism experiences the world — its niche is the sum of its experiences, rather than its immediate physical surroundings. Some plants, for example, can grow smaller or larger leaves, depending on whether they happen to be growing in a sunny or shady spot. So this is a form of niche construction, claims Sultan, because the plant is altering its own experience of sunlight.

Although phenotypic plasticity has been well studied by a number of researchers, it has yet to be incorporated into the core of evolutionary theory. "Niche construction weaves together a number of themes in ecology and evolution that have typically been studied in isolation," Sultan says. Rethinking evolution in light of plasticity and other issues raised by niche construction could contribute to an updating of evolutionary theory, Sultan suggests.

An update is precisely what Laland and his colleagues have proposed in what they have dubbed extended evolutionary theory. In classical theory, genetic inheritance is the only link through time between generations. Niche construction requires that a second form of inheritance, termed ecological inheritance, be taken into account.

Inherit the earth

According to this view, many of the physical features that a creature encounters, and the kinds of problem it has to solve, are inherited from the activities of the previous generation. Forest fires, for example, which help to distribute the seeds of some plant species, might be thought to rely solely on the chance of a lightning strike. But the plants in the forest can themselves increase the odds of a fire by secreting flammable oils and retaining dry dead wood in the canopy⁴. Similarly, every earthworm inherits an environment more suited to its lifestyle thanks to the activities of its forebears. Ecological inheritance means that the effects of genes on the environment are, a little like the genes, passed down through the generations.

The notion that genes reach beyond the bounds of the organism is often referred to as the 'extended phenotype', a term coined by

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Construction workers: humans create towns from deserts, but how do we and our niches interact?

Richard Dawkins, an evolutionary biologist at the University of Oxford, in his 1982 book of the same name. So it might come as something of a surprise that Dawkins has written a highly critical commentary accusing niche constructivists of a serious conceptual blunder⁵.

Dam fools

Dawkins's classic example of an extended phenotype is the beaver dam. These remarkable structures dramatically alter the surrounding ecosystem. Trees are felled to make the dam, which in turn floods the area, providing a new environment for species from frogs to fish. If the beaver's footprint on its environment is viewed as an example of ecological inheritance, it would seem that the extended phenotype and niche construction should make natural bedfellows.

But guess again. Although Dawkins says he recognizes the importance of organism-induced effects on the world, he believes that niche construction conflates two distinct kinds of effects. Dam-building certainly counts as an organism engineering its environment, he says, but other effects, such as the oxygenation of the atmosphere by cyanobacteria, are mere co-incident by-products of life. These types of effects, which Dawkins calls niche changing, are too loosely connected to the success of the organisms that cause them to count as genuine niche construction.

Dawkins is not alone in this view. Kim Sterelny, a philosopher of biology at the Victoria University of Wellington, New Zealand, says that niche construction "lumps too many things together". This matters, because the two kinds of effects, construction versus mere changing, generate different feedback loops between the organism and the environment,

which can lead to different evolutionary dynamics, Sterelny says.

Laland says he is sympathetic to the distinction, but is concerned that the term 'mere' associated with 'niche changing' downplays its evolutionary importance. For Laland, niche changing is as important to evolution as beaver-like niche construction. When you get down to doing the models it often doesn't help much to make the distinction, says Laland. The effects of organisms can have evolutionary consequences regardless of whether they are produced by adaptations.

Although the philosophical debates con-



Kevin Laland (left) thinks the power of niche construction is being underestimated, but Laurent Keller is not convinced.

tinue, other researchers are busily incorporating the ideas of niche construction into their work. Sultan, for instance, finds the concept useful in thinking about invasive species, whose potentially destructive power is a key issue in conservation biology.

Invasive species, such as weeds, often experience a time lag between arriving in a new niche and colonizing it. It may take a while for successful genetic variants of the invader to arise and spread, for instance. But if a species arrives that has sufficient phenotypic plasticity to thrive in the new environment, the take-over might be much more rapid. Sultan believes that explicitly adopting niche-constructivist views

on phenotypic plasticity could help scientists to devise appropriate strategies for combating conservation problems: it could give them, for example, more accurate tools for projecting the rate of spread of an invasive plant.

Others are pioneering ways to study perhaps the ultimate niche constructors — us. In many obvious ways, humans have utterly transformed otherwise inhospitable parts of the world to suit our needs, from ranks of houses in the desert to skyscrapers. Perhaps a less obvious example of niche construction is human culture. Culture itself can be seen as a niche that we inhabit, and just as we shape our culture, our culture shapes us. One example of this is the emergence over several thousand years of lactose tolerance in European adults, which has followed the cultural practice of drinking cow's milk⁶.

Culture club

Now a number of anthropologists are scrutinizing how culture can put selective pressure on our genetic make-up. In the past, many have been reluctant to tackle such questions, in part because of fears of being associated with genetic determinism, but also because of the daunting mathematics of modelling gene-culture interactions. But that seems to be changing, says Joe Henrich, an anthropologist at Emory University in Atlanta, Georgia. "The study of cultural evolution is expanding rapidly within scientific anthropology," he says.

One of the hottest areas at the moment is the puzzle of human sociality — why we are so often willing to cooperate with unrelated people, even when it is not in our immediate self-interest⁷. Whether or not genes promoting sociality flourish depends in part on the social environment in which they find themselves, which in turn is affected by culture. "We have shown that culture can evolve to change the selective environment faced by genes favouring cooperation. This opens up a whole evolutionary vista unavailable to non-cultural species," says Henrich.

Niche-construction advocates are passionate about their new view of ecological and evolutionary processes, whether they study bacteria or humans, but it is too soon to say whether the approach will yield insights that might otherwise have been missed. Still, Laland fully accepts the challenge. "The onus is on us to show that this is going to be useful," he says.

Dan Jones is a copy editor for *Nature Reviews Drug Discovery*.

1. Odling-Smee, J., Laland, K. & Feldman, M. *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, 2003).
2. Lewontin, R. C. in *Evolution From Molecules to Men* (ed. Bendall, D. S.) 273–285 (Cambridge Univ. Press, Cambridge, 1983).
3. Shachak, M., Jones, C. G. & Brand, S. *Adv. Geocol.* **28**, 37–50 (1995).
4. Schwilk, D. W. *Am. Nat.* **162**, 725–733 (2003).
5. Dawkins, R. *Biol. Phil.* **19**, 377–396 (2004).
6. Beja-Pereira, A. et al. *Nature Genet.* **35**, 311–313 (2003).
7. Hammerstein, P. (ed.) *Genetic and Cultural Evolution of Cooperation* (MIT Press, Cambridge, 2003).

J. MARSHALL/CORBIS

S. PRADA/UNICOM

Star of the south

This month South Africa will officially open the largest optical telescope in the Southern Hemisphere. But is the country ready to capitalize on its investment? **Michael Cherry** investigates.

Sutherland has a reputation as the coldest town in South Africa. Sat on a windswept, barren plateau at an altitude of some 1,600 metres, it is home largely to sheep farmers. But on 10 November, it is set to be the focus of a visit by President Thabo Mbeki, when he inaugurates the largest single optical telescope in the Southern Hemisphere at the town's nearby observatory.

When construction began on the \$32-million Southern African Large Telescope (SALT) five years ago, it was accompanied by a sense of excitement. "We are building a gigantic African eye through which we can view the Universe," Mbeki said at the time. For Sutherland, which is in South Africa's poorest province, the Northern Cape, the prospects seemed good. And the government had high hopes that big science projects such as SALT would inspire a future generation of South African scientists.

But even as the telescope saw first light on 1 September, some questions were being raised about South Africa's readiness for the



The Southern African Large Telescope has become the focus for training the country's budding astronomers.

project. "I'm most impressed with the engineering," says Brian Warner, emeritus astronomer at the University of Cape Town (UCT). "The embarrassment is that South Africa does not have the capacity to use the observation time allocated to it."

SALT is a collaboration between South Africa, Germany, New Zealand, Poland, Britain and the United States. Having contributed one-third of SALT's overall costs, South Africa has been given a similar proportion of observation time. But astronomy in South Africa — even more than other scientific disciplines — is still struggling to escape the confines of its colonial legacy.

That legacy has a long history. The plaque

outside the South African Astronomical Observatory (SAAO) in Cape Town declares that it was "established by the Lords Commissioners of the Admiralty in 1820 on the recommendation of the Board of Longitude". Yet administration of the SAAO, which has run Sutherland's observatory for the past 33 years, was transferred to South Africa as recently as 1971.

Since then, just one of its directors has been South African: Richard Woolley, who held the post from its inception until 1976. Even when South Africa's two top jobs in astronomy came up for grabs this year, they were taken by non-nationals. In July, astronomer Phil Charles at the University of Southampton, UK, became

SHARP SHOOTER

With its 11-metre hexagonal mirror, the Southern African Large Telescope (SALT) is both a copy and an upgrade of the Hobby-Eberly Telescope in Fort Davis, Texas. And it will complement its sibling by surveying stars and galaxies visible from the Southern Hemisphere.

One of the biggest advances on SALT is a system that aims to improve image quality over the Texas telescope. SALT features a spherical aberration corrector, which should provide sharper

images, a larger field of view, and it uses its entire mirror area for capturing starlight.

If the images taken by SALT when it saw first light in September are anything to go by (see right), the corrector seems to be a major success. Its designer, Darragh O'Donoghue, an astronomer at the South African Astronomical Observatory, says he couldn't have done it without help from the Texas team, who explained the problems with Hobby-Eberly. "When you start off on the first landing, to



This image of a galaxy some 30 million light years away was one of the first taken by SALT.

climb one more storey is that much easier," O'Donoghue says.

SALT will excel in studying astronomical objects that vary

over time, says its director Phil Charles. It will also make its mark in the ultraviolet spectrum, he adds, where the capabilities of its imaging spectrograph are unrivalled. This instrument, installed

in mid-October, will observe very faint objects, such as distant supernovae explosions that took place in the early Universe.

the new director of the SAAO and therefore of SALT. And earlier this year, Renée Kraan-Korteweg, a Dutch national, succeeded Warner in the chair at UCT — the only department in the country dedicated to astronomy.

Charles admits that the SAAO is run largely the same way it was a century ago. “It is outrageous that until recently there has been nothing in place to create an indigenous astronomy community,” he says.

The observatory’s track record in training graduate students has been particularly dismal, says Clifford Nxomani. For the past three years, Nxomani has run the SALT Science Foundation, an arm of the SAAO that aims to promote astronomy in the wider community. The approach of the astronomers at the observatory, he says, has been one of “focusing on their work, and regarding everything else, including student training, as an imposition”. Nxomani left the foundation last month to take up a senior post in South Africa’s Medical Research Council.

“The ten senior astronomers employed by the SAAO have only five doctoral students between them,” Nxomani notes. “As a national facility, one of its functions should be to train postgraduate students.” At UCT, it is a similar story. With three academic staff, its astronomy department, has produced just 15 PhDs since its inception in 1972. But Nxomani is optimistic that things will improve. “Everything is changing with the advent of SALT,” he says.

Scope for improvement

There have been some recent attempts to address the lack of indigenous astronomers. In 2003, the SAAO, under the leadership of astronomer Patricia Whitelock, helped to launch the National Astrophysics and Space Science Programme. Based at UCT, this master’s programme enlists the help of 46 scientists nationally who are active in astronomy, cosmology and space science to teach the classes. Students also go on field trips to SALT and other local observatories, before choosing a dissertation topic and supervisor for their thesis.

To date, 15 South Africans and 6 students from other African countries have graduated from the course. Another 19 South Africans and 5 students from the rest of Africa are currently enrolled. Peter Dunsby, a cosmologist at UCT who coordinates the programme, hopes that the students from other African countries will return home and form “nodes of expertise across the continent that would be linked through using SALT”.

The government’s Department of Science and Technology is currently considering a proposal to expand the programme to produce 76 doctoral graduates and fund 30 postdoctoral fellowships over the next five years. By providing training across many astronomy and space-



School children attend a science fair held at Sutherland’s observatory.

science topics, Dunsby hopes to generate graduates not only for SALT, but also for other fields of space science, such as radio astronomy. That strategy could pay off as South Africa is trying to expand its role in space science. It is currently bidding against China, Australia and Argentina to be the main site of the US\$1.2-billion Square Kilometre Array radio telescope, which will be decided in mid-2007.

The success of that bid may depend on how SALT is perceived — especially at home. Local investment in SALT has been justified on several grounds. South Africa has a tradition in

observations in bed. Although this system is more flexible than traditional scheduling, some worry that it might perpetuate astronomy’s colonial legacy as it will severely limit the number of scientists visiting the observatory.

In addition, four of SALT’s team of six dedicated astronomers have so far been appointed, and none of them is South African. “The problem,” says Nxomani, “is that the job requires experience in working on large telescopes. Only now that we have SALT, can South Africa begin to develop this capacity.” Whitelock confirms that there have been no South African applicants with the necessary experience.

Seeing how slowly things change at the top, Nxomani chose to focus his efforts at ground level, working with local schools in Sutherland. He got pupils to take part in educational camps in Cape Town and at SALT. Schools from all over the country are also encouraged to visit both sites. There has been a similar push to improve the skills of teachers, many of whom have attended an 18-month part-time course to upgrade their teaching abilities in maths, science and technology.

Teething troubles

The problem with astronomy in South Africa, as in so many other areas, originates in the education system. The post-apartheid government has been fairly unsuccessful at improving the maths and science education of black Africans since it took office in 1994. In 2002, for example, the number of black school-leavers who had maths grades good enough to study science at university was the same as it was in 1991, according to a report published last year by the Johannesburg-based Centre for Development and Enterprise.

So far, the country as a whole has produced only three black and two mixed race PhD astronomers. But retention rates have been high: four of them are now at the SAAO; the other is at the University of Chicago.

Nxomani agrees that the problem starts at grassroots level. He recalls his shock at the low standard of facilities at Sutherland High School, especially as during the apartheid era it was a white school. “In the Northern Cape, everyone is poor — whether you are black or white,” he says.

Despite these problems, the advent of SALT has given a new lease of life to astronomy in South Africa. “The future is bright, because if we use our observation time wisely, we should be able to make a significant impact in astrophysics and cosmology,” says Dunsby. “The challenge is to ensure that sufficient local astronomers are trained to take advantage of this opportunity.”

Michael Cherry is Nature’s contributing correspondent in South Africa.

“One of the South African Astronomical Observatory’s functions should be to train postgraduate students.”
— Clifford Nxomani



optical astronomy that dates back to the early nineteenth century — the British astronomer John Herschel, for instance, came to Cape Town in the 1830s to map the southern skies, and while there documented the return of comet Halley in 1835. And 60% of SALT’s construction and development budget has been spent in South Africa — resulting in a net inflow of capital — although only a fraction of this was spent in the poor Northern Cape.

Like its sister telescope in Texas (see ‘Sharp shooter’, opposite), SALT will be run more like a space telescope than one based on the ground. Astronomers will submit requests for observations over the Internet, and once these have been prioritized and completed by operations staff at SALT, they will receive their data over the net too. Warner quips that in his retirement he will be able to do his

Winds of change

Hurricanes can grow more intense in a matter of hours, but exactly why remains a mystery. **Mark Schrope** flies into the eye of a storm to investigate.

In the space of one October day, Hurricane Wilma escalated from a tropical storm to become the strongest hurricane from the Atlantic basin on record. It was a shift that no one saw coming, and Mexico's Yucatán Peninsula bore the unexpectedly strong brunt of the storm.

Over the past few decades, hurricane forecasters have dramatically improved their predictions of the paths that storms might follow. The accuracy of such 'track predictions' made 72 hours before a hurricane reaches land, for instance, have improved by more than 50% since 1970, according to the US National Oceanic and Atmospheric Administration (NOAA). But over the same period, models for hurricane intensity have remained relatively primitive. The turbulence that drives the internal dynamics of a hurricane is mind-bogglingly complex. And without understanding that turbulence, and its effects on a storm's intensity, forecasters haven't really been able to improve their overall long-range forecasts.

This year, correcting that situation became a high priority for hurricane research centres across the United States. During the most active Atlantic hurricane season on record, several new research programmes have wrestled with the question of intensity. The move may come not a moment too soon: two recent studies have suggested that hurricanes might grow more intense in the future because of climate change^{1,2}. If so, the millions of people who live in vulnerable coastal areas will be at greater risk than ever, further underscoring the importance of good forecasts.

"If you have accurate track and intensity forecasts that people can rely on, it will have a huge societal benefit," says Richard Anthes, a hurricane specialist and president of the University Corporation for Atmospheric Research in Boulder, Colorado. "You would save enormous amounts of energy and protect huge numbers of lives."

One of the most extensive new programmes is a three-year, US\$3-million plan to investigate how, why and when a hurricane's intensity changes. Funded by the National Science Foundation, the Hurricane Rainband and



In a spin: understanding what drives changes in the intensity of a storm means gathering data from the heart of hurricanes.

Intensity Change Experiment (RAINEX) combines a massive set of plane observations with experimental models. During August and September, RAINEX coordinated three aircraft equipped with Doppler radar to fly into the hearts of storms. All of the planes were WP-3D Orion 'hurricane hunters', more commonly known as P-3s. Two of them were run by NOAA; the third was owned by the Navy and carried ELDORA, the most advanced Doppler radar currently on hurricane duty.

Whirlwind start

Four years ago, when researchers first proposed RAINEX, they had no idea that their first year of action would bring them a nearly ideal set of storms. As coastal residents were battered by hurricane after hurricane, the RAINEX team gathered a bounty of data that may one day help protect against such devastation.

In August, Hurricane Katrina tantalized forecasters as its bands of rain organized themselves symmetrically around the storm's eye, and then failed to develop into a feature that could have weakened the storm's intensity. In September, Hurricane Ophelia spun

seemingly endlessly along the US southeastern coast, allowing the team to observe how a storm maintains its intensity over relatively cold water. Later that month, Hurricane Rita suddenly blew up in strength to reach the highest, category 5, rating.

Aboard one of the P-3 flights into the heart of Rita after it had subsided to category 2, Brad Smull from the University of Washington in Seattle described the hurricane's unpredictability. "During an eight-hour flight, the storm went from category 3 to category 5," he said over the din of the plane's engines and the hurricane. "That's very exciting." Repeatedly flying through the hurricane to reach Rita's eye brought its own kind of excitement, and it was difficult for me, a first-timer on the plane, to appreciate how relatively weak the storm had become. At times I had to clutch the handrail running the length of the P-3's ceiling to steady myself as the plane bucked.

One of the hypotheses that RAINEX is testing is the idea that hurricanes grow in intensity because the bands of rain that swirl around the storm inject energy into the storm's eye. When the winds immediately around the eye —

J. SCHMALTZ/MODIS/NASA GSFC

known as the eyewall — are strengthening, rainbands tend to spiral into the eye. At other times, the outer rainbands can begin to form their own circle, or secondary eyewall, around the eye. If this happens, the secondary eyewall often starves the inner eye of energy so that it begins to fall apart. And in some cases the secondary eyewall lasts long enough to become the primary eyewall, and even begins to strengthen further.

Knowing exactly when and how this eyewall replacement happens is important for hurricane forecasts. But finding out means studying the flow of energy within a storm, which is no simple task.

Hurricanes are fuelled when evaporation transfers heat from the ocean into the atmosphere. The resulting water vapour releases energy as it rises and condenses into clouds; this process heats the surrounding air, causing it, too, to rise. All this upward movement of air causes the pressure below it to drop and winds to start blowing.

Surprisingly, estimates suggest that there is little difference between the amount of energy transferred from the ocean to a tropical storm and to a category 5 hurricane. But scientists are hard-pressed to explain why, in a hurricane, much of the transfer is concentrated in the relatively small area of the eye. These zones of energy transfer are crucial to controlling a storm's shifts in intensity, but are hard to study because they can be small and move around rapidly.

Flight coordinators

Getting a good picture of how the eyewall and rain bands interact requires observing both structures simultaneously. But typical hurricane reconnaissance flights can't do that because they mainly crisscross a storm's eye. The RAINEX plan uses three planes: one flying through the eye, one flying on the inside edge of the main rainband, and the other taking the outside edge of the band — something that had never been done before. "It's really fantastic that we have been able to accomplish this," says Smull.

Coordinating this aeronautical dance wasn't easy. The planes maintained a satellite connection to a computer chat room, so that scientists in the air could 'talk' to each other and to the command centre in Miami, Florida. Flight paths were constantly updated based on radar data as well as the chat-room information. "It was a very nerve-wracking process," says Shuyi Chen of the University of Miami, who with Robert Houze of the University of Washington is principal investigator for RAINEX.

Each plane uses Doppler radar to gather core data about cloud structure and winds, which reveal locations of greatest turbulence. But expanding the radar's information into a three-dimensional view of the storm requires



Researchers collect data (above) as they fly through Hurricane Rita onboard a P-3 plane (below).



vertical profiles of temperature and wind. The planes can't fly straight down to acquire that, nor can they fly safely at levels below 5,000 feet, so they use instruments called dropsondes. At key points the plane releases dropsondes, which radio back real-time information on temperature, pressure and wind speed until they hit the water.

But improving understanding of hurricane intensity takes more than data from planes. "Observation doesn't really give you the full picture," says Wen-Chau Lee, a RAINEX investigator at the National Center for Atmospheric Research in Boulder. "Even with eight-hour coordinated flights, we're still getting only snapshots, and we don't sample the storm for the entire 24 hours."

Whipping up a storm

Making sense out of these snapshots requires computer modelling. At the operations centre in Miami, researchers use experimental models to compare their predictions with data arriving from the hurricane flights.

If scientists could better understand why hurricanes change intensity, they might also be able to improve their predictions of a storm's path, says Lee. Large-scale environmental

factors play the biggest role in influencing where a hurricane goes, but internal dynamics are also important.

Along with RAINEX, various other projects under way this year will help hurricane researchers to answer the many questions about intensity changes. Several fall under the umbrella of the Intensity Forecasting Experiment (IFEX), which is managed by NOAA's Hurricane Research Division in Miami.

IFEX covers a variety of projects, including studies of tropical cyclones forming in their earliest stages in the eastern Pacific, as well as storms decaying in the northern Pacific. Historically, researchers have tended to study only mature storms, leading to deficiencies in understanding the complete storm cycle. Another IFEX programme has been testing unmanned aircraft for monitoring the lower levels of storms at resolutions much higher than is possible with dropsondes.

Many hurricane experts are confident that all this work will soon pay off. Anthes, for one, says that better observational and satellite data, combined with improved computers, should help to improve intensity forecasts in coming years. "I see no fundamental reason why we shouldn't be able to predict intensity much better," he says. "We just need very high-resolution models and good understanding of cloud physics and dynamics."

One such high-resolution tool is already under development at NOAA's Environmental Modeling Center in Camp Springs, Maryland. Called the Hurricane Weather Research and Forecasting model, it is scheduled to begin operation in 2007. It will incorporate the conclusions of RAINEX, IFEX and other ongoing projects, and will have nearly an order of magnitude better resolution than the main models currently in use. For coastal residents who are wary of the coming hurricane seasons, such improvements can't come soon enough. ■

Mark Schrope is a freelance writer in Florida.

"I see no fundamental reason why we shouldn't be able to predict hurricane intensity much better."
— Richard Anthes

1. Emanuel, K. *Nature* **436**, 686–688 (2005).
2. Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H.-R. *Science* **309**, 1844–1846 (2005).

BUSINESS

M. TAMURA/GETTY IMAGES

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Giving it their best shot: vaccine producers are focusing their efforts on tackling flu viruses.

Race is on for flu vaccine

Drug companies are using adjuvants to boost their vaccines in a bid to be ready for a flu pandemic, as **Meredith Wadman** reports.

Not so long ago, vaccine manufacture was a neglected backwater of the drug industry. The threat of a global flu pandemic has changed that. Leading vaccine producers are now engaged in a furious race to be ready to tackle such a pandemic.

They face a considerable challenge. The cell-surface proteins of flu viruses change, or 'drift', over time, so vaccine makers can't develop a vaccine in advance that they know will work if bird flu acquires the ability to pass between humans, triggering a pandemic. What's more, there aren't many companies in the race, after low profits drove many to stop producing vaccines at all.

It has been estimated that today's combined global manufacturing capacity could provide a vaccine for 450 million people at most (D. S. Fedson *J. Public Health Policy* **26**, 4–29; 2005). And that calculation might be optimistic: it assumes that two vaccinations of 15 micrograms each would confer protection, whereas one recent trial suggested that two doses of 90 micrograms would be required.

"The biggest challenge unequivocally is vaccine production capacity," says Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases (NIAID).

Because of this capacity issue, vaccine producers are looking hard at ingredients known as adjuvants, which render vaccines more effective at low doses. France's Sanofi-Pasteur and CSL in Melbourne, Australia, are already conducting trials of candidate pandemic vac-

cines that use adjuvants made of alum, an aluminium salt — the only kind already approved for use in humans in the United States.

Sanofi-Pasteur's results from a trial of 300 subjects comparing vaccine alone with a vaccine containing an alum adjuvant, at a dose as low as 7.5 micrograms, are expected by the end of the year. And GlaxoSmithKline is planning to launch its own trials of a vaccine containing an alum adjuvant early next year.

But it is California-based Chiron — which is currently being taken over by Novartis — that may have the most promising adjuvant. Called MF59, it consists of emulsified squalene with influenza virus in the centre of each droplet. In clinical trials of an early candidate vaccine aiming to confer protection against H5N1 — the virus that has been killing birds by the million in Asia and Europe — vaccine containing the MF59 adjuvant was significantly better than vaccine alone at spurring the production of antibodies to H5N1 (K. G. Nicholson *et al. Lancet* **357**, 1937–1943; 2001).

Coping with change

Perhaps the most exciting thing about MF59 is its apparent ability to confer protection against H5N1 even if the virus's cell-surface proteins change. The antibodies induced in the blood of volunteers immunized against the H5N1 virus that infected Hong Kong in 1997 were later found to neutralize versions of the virus that appeared in southeast Asia in 2003 and 2004 — even though they had

different cell-surface proteins (I. Stephenson *et al. J. Infect. Dis.* **191**, 1210–1215; 2005).

"We at Chiron believe we have the only practical answer to vaccination against pandemic flu," says Rino Rappuoli, the company's chief scientific officer. "The adjuvant allows you to vaccinate and to cover strains that may be drifting," he says. "So you can actually go from a strategy of containing a pandemic to preventing a pandemic completely by vaccinating people before it comes."

But the ability of MF59 to protect against a drifted strain of H5N1 cannot be proven until a pandemic strain actually emerges. "It depends on how great the drift is," says Jerald Sadoff, president of the Aeras Global TB Vaccine Foundation in Bethesda, Maryland. "And since we don't know that, it's not clear whether it would work."

The mechanism that allows adjuvants to prompt a more vigorous immune response isn't well understood. But they are thought to act by recruiting and activating the immune-system cells that respond to vaccine proteins, and by prolonging the time that a vaccine's active ingredient is exposed to the immune system.

Seeking approval

Alum adjuvants have a good safety record, and MF59 has been used in flu vaccines in Europe since 1997. But new combinations of adjuvants and vaccines can raise fresh safety issues — and it isn't clear how quickly MF59 will be approved in the United States.

Geoffrey Porges, an analyst at Sanford Bernstein, a New York investment bank, thinks that Chiron could face significant regulatory hurdles in bringing a flu vaccine containing the MF59 adjuvant to the US market: "It stands a better shot of getting to market quickly in Europe," he says. "In the United States, regulators might be wary of the new vaccine and new adjuvant combination, and more traditional vaccine approaches could have an advantage."

MF59 has also yet to be proven against H5N1 in large clinical trials. NIAID will shortly test Chiron's H5N1 vaccine in some 400 subjects, Fauci says, with another trial of Sanofi-Pasteur's alum-adjuvant vaccine to follow next year. Both trials will test vaccines formulated with adjuvant at the time of manufacture against vaccines in which adjuvant is added just before injection. If the latter is effective, existing vaccine stockpiles could be made to go further with the adjuvant's help.

Much is likely to depend on the results of the upcoming trial of Chiron's vaccine. "We'll have to work with the Food and Drug Administration to get enough data to see if we can get it licensed," Fauci says. "And that's going to depend purely on the data."

UK must go on promoting and funding science

SIR — Your Editorial “Blair’s failure” (*Nature* 435, 129; 2005) refers to a “declining interest amongst the young in science as a career” and states that the UK government should either abandon its target of raising research and development spending from 1.9% to 2.5% by 2014, or explain it. I believe that the UK government must retain this target, as the science and engineering base is vital to future global competitiveness. For example, many wealth-creating, innovative companies will develop from Britain’s world-class small high-tech companies.

If the UK government is to achieve its economic and social ambitions, it must maintain a strong, diverse supply of scientists to sustain the research base. The brightest and most creative young people need to be inspired to take up careers in science, engineering and technology. Indeed, the number of full-time undergraduates studying for first degrees in science and related subjects is actually rising. And although the media portrayed the closure of two UK university chemistry departments as a major failure, such decisions are influenced by a range of factors, including student demand and a move to concentrate research into larger departments with higher ratings.

There are difficulties in comparing the statistics from the UK Higher Education Statistics Agency, as a result of changes in the classification of subjects from 2002–2003 onwards. However, between 1997–1998 and 2003–2004, the number of people studying science subjects in degree courses increased by more than 30%, at a time when there was a smaller increase (about 20%) in numbers studying for first degrees overall. The largest increases over this period have been in the biological (including psychology and agriculture) and computer sciences. The rise in popularity of computer sciences is at least in part attributable to the perception of career opportunities. The increase in the numbers studying medicine and dentistry has been accompanied by a welcome increase in the participation of women.

The picture for engineering, technology and the physical sciences is more complex, with subjects such as civil engineering and chemistry attracting fewer students while others, such as astronomy and aeronautical engineering, become more popular. Overseas students account for an increasing proportion in areas such as electronic and electrical engineering, presenting opportunities to retain the best.

Overall, the proportion of students studying for UK degrees in the sciences increased from 38% to 41% between 1997–1998 and 2003–2004, a reality very different from the picture painted in the

media. Of course there is no reason for complacency: A-level (17-year-old) entries in mathematics, computer sciences, physics, chemistry and biology averaged a decrease of 7.5% between 1997–1998 and 2003–2004. The UK government needs to continue to take action to enthuse people about the benefits of science education and the potential offered by careers in science.

Such government measures include ‘golden hellos’ to new science and mathematics teachers, a science and engineering ambassador scheme (in which some 8,500 young scientists act as role models for school students) and investment, jointly with the Wellcome Trust, in Science Learning Centres.

A recent MORI poll (see www.mori.com/polls/2004/pdf/ost.pdf) reported that 85% of people believe science makes a good contribution to society. I challenge the media to reflect the positive view of science and technology held by the majority of people in Britain today.

David A. King

Office of Science and Technology,
1 Victoria Street, London SW1H 0ET, UK

Universal fungus register offers pattern for zoology

SIR — The proposal by Andrew Polaszek and colleagues in their Commentary article (*Nature* 437, 477; 2005) for a universal register for animal names as a requirement of the next edition of the International Code of Zoological Nomenclature is very welcome. The authors note that a system was tried out by botanists and approved in principle by the International Botanical Congress in Tokyo in 1993, but was rejected by the subsequent congress in St Louis, Missouri, in 1999 and never implemented.

However, a formal system of registration of names was first proposed by mycologists 50 years ago (G. C. Ainsworth & R. Ciferri, *Taxon* 4, 3–6; 1955). Many mycologists were disappointed at the St Louis rejection, and in response, they established MycoBank in 2004 (P. D. Crous *et al.* *Mycol. Res.* 108, 1236–1238; 2004). Authors can obtain a unique accession number for their new entities through the web from MycoBank.org before publication, much in the way GenBank accession numbers are requested. New descriptions with MycoBank numbers are already appearing in the fungal literature.

The system is voluntary at present, but it is anticipated that leading journals, now starting to recommend this practice, will make it a requirement for the acceptance of papers with new scientific names in them as it becomes better known — just as they already require GenBank numbers for molecular sequences.

MycoBank, led by the Centraalbureau voor

Schimmelcultures fungal biodiversity centre in Utrecht, is being developed in close collaboration with Index Fungorum, an online database of 380,000 scientific names. MycoBank accession numbers are also being incorporated into the life science identifiers scheme.

MycoBank is sure to evolve and become more honed as use of it increases. It provides a model that zoologists and others interested in bringing increased order into the naming of life on Earth might examine.

David L. Hawksworth

IUBS/IUMS International Committee on Bionomenclature, Departamento de Biología Vegetal II, Universidad Complutense de Madrid, Plaza Ramón y Cajal, Madrid 28040, Spain

Mapping the complexities of science and politics

Your Editorial (*Nature* 436, 152; 2005) related the story of how *Nature* had to carry out the painstaking task of rounding up and destroying several thousand copies of last November’s supplement “China voices II” containing a map of China without Taiwan, then reprinting it with no map. You assured readers that “whether Taiwan is an independent country ... is not an issue on which *Nature* takes a stand”.

Science recently made a similar clarification — that it has no policy on the Taiwan question — in a published apology by editor-in-chief Donald Kennedy (*Science* 309, 1677; 2005), following that journal’s publication in July of a map of China that included Mongolia but not Hainan and Taiwan.

It is a sad reality that people in some quarters in China might still believe that this omission by two leading science journals represent elements of a concerted move with hidden political motivation, no matter what actions were taken by each journal to claim otherwise.

On the other hand, declaring whether to take a stand on a political issue that has nothing to do with science is a significant political statement in itself, especially when it comes from journals with the stature of *Nature* and *Science*.

It is a sad day for science and scientists everywhere when these journals are compelled to explain themselves by making declarations of this kind.

Ying-Hen Hsieh

Department of Applied Mathematics, National Chung Hsing University, Taichung, 402 Taiwan

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

BOOKS & ARTS

Scientists on screen

Does Hollywood think we're all dangerous megalomaniacs with crazy hair?

Mad, Bad and Dangerous: The Scientist and the Cinema

by Christopher Frayling

Reaktion Books: 2005. 239 pp. £19.95, \$35

Adam Rutherford**Joey:** It's like reading a script. Like, "This is a *Tyrannosaurus rex*, a creature from the Jurassic period."**Ross:** Actually, Joey, it's the Cretaceous period.**Joey:** Yeah, but, I can pronounce Jurassic.

Watching *Friends*, the pre-eminent US sitcom of the 1990s, I once found myself in the embarrassingly geeky position of correcting Joey's gaffe half a second before Ross did. Fortunately, I have few friends myself and this smart-arse moment passed almost unnoticed. Ross is an urban New Yorker, a loving father with a tight network of close friends. He is attractive and socially well adjusted (if a bit wet). He is also a lab-based palaeontologist, a postdoc working in a prominent metropolitan museum.

Ross spends little time writing grants and a lot of time hanging around coffee shops. In this regard, his is a hugely inaccurate portrayal of a working scientist. But at least he doesn't display any of the stereotyped traits associated with scientists in the movies, outlined in the title of Christopher Frayling's new book, *Mad, Bad and Dangerous*.

Frayling has reviewed in detail the changing role and perception of scientists in more than a hundred years of cinema. His analysis is partly chronological but desperately uneven. Frayling, a cultural commentator with a clear love of cinema, frequently comes across as a fanboy, albeit a rather academic one.

He covers films from the first half of the twentieth century rigorously, detailing lost or forgotten reels with the precision and loving hand of a devoted film historian. But his coverage of the modern era is more sporadic. The rather desultory selection of films he analyses in detail include some that might not figure in a cinéaste's list of classics at all. A detailed deconstruction of Stanley Kubrick's cold-war masterpiece *Dr Strangelove* is welcome after tracts on many films from the silent era. But to my mind, *The Man with Two Brains* falls into the category of films, like *Blazing Saddles*, that are remembered as being funnier than they actually are. It's a clumsy comedy, slapstick

funny at best, and doesn't strictly feature a scientist (Steve Martin's Dr Hfuhruhurr is a brain surgeon), let alone a particularly stereotyped version of one. Its presence in this book seems unwarranted.

Devoting more than a page to mercifully forgotten Australian funny man Yahoo Serious and his frankly dreadful film *Young Einstein*, while only mentioning in passing the original 1972 version of *Solaris* (Steven Soderbergh's remake doesn't get a look in), is at best an idiosyncratic decision. And where is *The Day After Tomorrow* (2004), the mega-budget eco-disaster film that, thanks to some canny marketing, received much coverage in the scientific press, this journal included?

It doesn't bother me that Penelope Ann Miller wears a cocktail dress to do battle with the mutant beast, half reptile, half insect, half human (it's that bad), in updated B-movie *The Relic*, as Frayling points out in one of the few sections detailing the portrayal of female scientists. She was, after all, at a swanky party when the beast got hungry. I was much more impressed when seeing this as a genetics undergraduate with the (fairly) accurate use of the enzyme reverse transcriptase to explain away the genesis of this wicked chimaera. Maybe Joan Allen's sensitive and passionate stem-cell biologist from Sally Potter's film *Yes* was too recent to be included.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Just a typical scientist?
Dr Strangelove, played by
Peter Sellers, does battle
with his uncontrollable arm.

Frayling has certainly done his homework, and *Mad, Bad and Dangerous* is bulging with factoids. Did you know that scientists feature as heroes in only 11% of horror films, but that mad scientists have caused bad things to happen in an eye-popping 31%? I found myself scratching my (egg)head at these data, but mostly in bewilderment at their relevance. Sometimes it's not clear whether he's talking about scientists, doctors, psychiatrists or explorers, and frequent references to television programmes obscure the line between the small and big screens.

Several studies exploring the stereotyped image of a scientist come under Frayling's gaze. These are interesting enough, if not wholly surprising. The influence of Hollywood imagery cannot be underestimated. When mentioning Frankenstein, we tend to think of Boris Karloff's flat-headed creature with a bolt in its neck, although this depiction belongs entirely to Universal Pictures and not Mary Shelley. But it seems that Einstein's late career decision not to comb his hair much was just as significant in creating a stereotype that scientists are invariably detached, bespectacled, crazy-haired social misfits.

Frayling's implicit conclusion is that scientists have been portrayed inaccurately throughout the history of cinema, although their dual role as unlikely hero and evil baddy has shifted

a little over time. Peter Sellers' Dr Strangelove is described as the archetype, certainly without morals, but where's the crazy hair? Do we as cinema-goers (or indeed as scientists) turn to Hollywood for accurate portrayals of scientists? Of course not. Nor should we, any more than we expect accurate portrayals of politicians,

pirates, spies, drunks or even penguins in the movies (and that includes the 2005 surprise hit documentary *March of the Penguins*). Perhaps we should praise *Friends* for portraying a scientist as someone who is just a little bit normal. ■ Adam Rutherford is an editor at Nature Publishing Group.

Return to the fortress

The Science and Fiction of Autism

by Laura Schreibman

Harvard University Press: 2005. 304 pp.
\$27.95, £17.95, €25.80

Michael Fitzpatrick

When my son was diagnosed as autistic some ten years ago, a friend's mother — a retired social worker — lent me a battered copy of Bruno Bettelheim's book *The Empty Fortress* (Free Press, 1967). This was the book that popularized the psychodynamic theory that autism was a result of defective parenting — a notion that caused much distress when children were removed from their 'refrigerator mothers' and 'aloof and obsessive fathers', and subjected to intensive psychotherapy in residential institutions. By the time I read Bettelheim, these theories had long been discredited and his book was of merely historical interest.

In *The Science and Fiction of Autism*, Laura Schreibman, professor of psychology at the University of California, San Diego, provides a comprehensive account of controversies in the field of autism. She discusses the impact that psychodynamic theories of autism had in the 1960s and 1970s despite the lack of scientific evidence supporting either them or the therapies with which they were associated. She recalls that, in the 1980s, the mother of an autistic child removed Bettelheim's books from the local library and threw them away, and mobilized other parents to do the same when they were replaced. Though I balk at any form of book-burning, parental anger was understandable given the consequences of Bettelheim's theories. As Schreibman acknowledges, the legacy of blaming parents has left a lingering "atmosphere of distrust and suspicion" between some parents and professionals in the sphere of autism.

With more than 30 years of clinical experience, Schreibman brings a valuable historical perspective to her discussion of the controversies surrounding the diagnosis of autism, its causes and its treatments. This is particularly helpful to parents, who may turn in distress and desperation to the latest wonder treatment. From this book they can learn how an earlier generation of parents was drawn, with familiar promises of recovery and cure, down the dead ends of 'holding therapy' or 'facilitated communication'. As Schreibman observes: "One need not be a scientist in order to know

how to evaluate information critically; one just needs to be appropriately critical." Her opening chapter provides a readily accessible guide on critical evaluation that could save parents — and their children — from much grief.

Schreibman's experience is also apparent in her survey of her own sphere of expertise: behavioural treatment programmes. She presents a detailed critique of some of the extravagant claims made on behalf of Ivar Lovaas'

and dietary therapies). Working in alliance with rogue scientists and doctors, groups of parents have linked up through the Internet to pursue strident campaigns promoting interventions of unproven efficacy and uncertain safety. Like earlier campaigns based on miracle cures, these are likely to prove damaging to the families of autistic children, but on a wider scale. If these campaigns undermine child immunization programmes, other families too will be damaged — in Britain we have already seen outbreaks of measles and cases of measles encephalitis in unvaccinated children.

One largely unrecognized aspect of Bettelheim's legacy is that professionals in the world of autism, burdened by the guilt of the past, tend to be reticent in confronting the junk science that underlies much contemporary parent activism. Although today's activist parents never suffered from psychodynamic theories as an earlier generation did, they do not hesitate to cite Bettelheim to legitimize their sense of victimhood and their rage against the medical



Child prodigy Matt Savage uses jazz to escape the isolating effects of autism. But how do parents cope?

'applied behavioural analysis', while noting that, for some children, early behavioural interventions have shown impressive results. Her approach is refreshingly undogmatic and pragmatic, emphasizing the importance of applying theory flexibly in relation to the particular needs and circumstances of the child. She is impatient at polemic over 'inclusion', dismissing the controversy over whether autistic children should be taught at home or at school as "another meaningless debate", insisting that what matters is how they are taught, not where.

In her concern to mend fences on the troubled boundary between parents and professionals, Schreibman perhaps underestimates the challenge arising from the current wave of parent activism around "unorthodox" biomedical theories (such as links between autism and vaccines) and treatments (such as chelation

establishment. The result is that activist groups, which are unrepresentative of parents in general and not accountable to them, and which often have links with commercial interests that provide expensive investigations and treatments, have enjoyed growing influence without facing a concerted challenge from experts in the relevant disciplines. If Schreibman's book encourages both parents and professionals to adopt a more critical approach towards such campaigns, this will protect families from further "crushed hopes, ineffective treatments and false starts". It will also help to restore appropriate boundaries between parents and professionals. It may therefore make an important contribution to the welfare of children in general, and those with autism in particular. ■

Michael Fitzpatrick is a GP in London and author of *MMR and Autism: What Parents Need to Know*.

Green in tooth and claw

Demons in Eden: The Paradox of Plant Diversity

by Jonathan Silvertown

University of Chicago Press: 2005. 192 pp.

\$25, £17.50

Peter D. Moore

Apart from a few parasites and decomposers, all plants face the same problem: how to capture light energy and fix atmospheric carbon dioxide more effectively than their neighbours. Actually, a darwinian would argue that their ultimate problem is how to leave more progeny than their neighbours, and that photosynthetic competition is just one aspect of this struggle. Growth may help, but pollination, seed maturation, fruit dispersal and seedling establishment all contribute to the final outcome in the next generation.

The theory of natural selection is based on the idea that heritable variation, acted on by the pressures of the environment, ultimately favours excellence. Only the fit survive. This being the case, one might predict that one supremely fit plant species, capable of photosynthesis, vegetative growth, pollination and seed dispersal in a way that outperformed all other plants, might ultimately prove the victor in the evolutionary struggle and would dominate the world's primary production. But this has not happened.

On the contrary, Earth displays an extra-

ordinary diversity of plant species (around 400,000), almost all of which contribute to the planet's energy-trapping potential. Why? How can biodiversity be reconciled with natural selection? This is the question Jonathan Silvertown asks in *Demons in Eden* as he explores the remarkable variations of form and the elaborate systems of coevolution with animals that are found throughout the plant kingdom.

Among plants there are some darwinian demons that seem unstoppable in their spread, such as the kudzu vine, which seems intent on burying Florida, or the rhododendron, which is running amok in British woodlands. But both of these were introduced by humans, taken to parts of the world where there seem to be no constraints upon their spread. And herein seems to lie the answer. Constraints on population expansion decelerate the tendency for any plant species to spread indefinitely and become the ultimate demon. Such constraints include herbivores, seed predators, parasites, diseases and new evolutionary developments in competitor plants.

When a plant escapes from these constraints, it may for a while become a darwinian demon. For example, genetic evidence seems to show that the olive tree invaded the Canary Islands just once. Freed from predators and competitors it may well have spread unrestrained. There is still nothing on the Canaries that eats olive seeds, but the arrival of other

tree species and the introduction of goats by people supplied the constraints that put the demon olive back in chains.

Another type of constraint on indefinite population expansion in plants is the supply of resources. Soil elements, such as nitrogen, phosphorus, potassium and calcium, are not always abundant, and the one that falls farthest behind the plant's demands may become the limiting factor for its growth. Silvertown describes the long-running Park Grass experiment at Rothamsted in England, and shows how the addition of an element, such as nitrogen, that may be limiting to some plant species, causes changes in the composition of grassland, releasing some species from the constraints of element limitation and causing others to be suppressed by the growth of the unchained demons.

The great diversity among plants does not therefore contradict darwinian principles, but upholds them. Vegetation illustrates the complex balance of interacting forces and trade-offs that evolution generates among competing species. In this highly readable and pleasantly anecdotal account of the dynamics of the plant world, Silvertown suggests that tasting the fruit of evolutionary knowledge may provide us with a ticket for readmission to the Garden of Eden, where we can exercise the privilege by ensuring that biodiversity is conserved. Let us hope that he is right.

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK.

EXHIBITION

In the croak room

Nearly one-third of the world's amphibian species are endangered, and countless have already been lost. Frogs used to be creatures of the wild, but are fast becoming creatures of the lab — pickled, jarred and preserved as a static piece of history.

Now some long-dead frogs are taking centre stage as part of an exhibition at the University of Kansas. Creatures from the university's herpetology collection, along with some from the Field Museum in Chicago, have been cast in urethane in a floor-to-ceiling display of froggy glory.

The show was created by Tracy Hicks, an artist from Texas who has been fascinated by natural-history collections since childhood, when he hatched eggs from turtles, lizards and snakes in his bedroom. The new installation is his tribute to vanishing species and to the science and art of collecting.

Floating in more than 1,300 jars are casts of some 79 species of frogs and tadpoles, many of them extinct, others endangered. They include *Atelopus ignescens* — once so common in the Andes that they would crunch frequently and



unpleasantly under the tyres of passing cars — and the extinct golden toad of Costa Rica's cloud forests. Other featured species are just bizarre, such as the flying frogs from Asia or the paradox frog, in which a tadpole 25 cm long metamorphoses into an 8-cm frog.

In several cases, Hicks made his casts from the species holotype. He practised for years on less valuable specimens to perfect his technique, and both museums, in Chicago and

when an ultraviolet light is switched on, transforming the wall of jars into an eerie spectacle, where details such as skin impressions pop out, glowing blue or green. The effect communicates the life of the animals. Audio recordings add to the impression, with a frog chorus of now-vanished species echoing their chants along the walls.

The exhibit, 'Two Cultures: Collection', will remain on display until March 2006. A.W.

Kansas, eventually gave him permission to work with their priceless specimens. "This was an opportunity to show that the scientific collection has an aesthetic value to it, a certain beauty," says John Simmons, manager of the university's collection and a collaborator on the exhibit.

Most strikingly, Hicks' translucent casts fluoresce

TRACY HICKS

Wit and wisdom

From pioneering xerographer to innovative teacher, Georg Christoph Lichtenberg was a physicist with many skills, but perhaps most remembered will be his acerbic aphorisms.

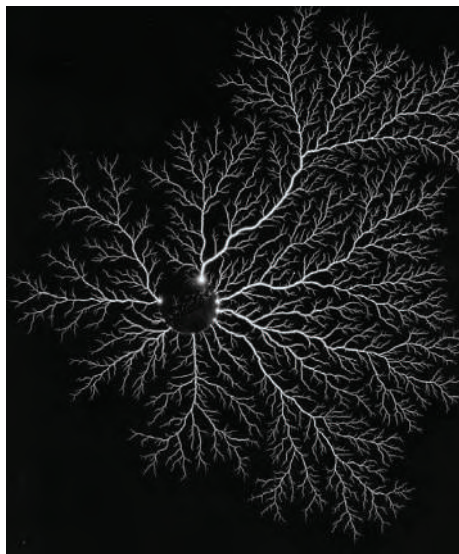
John L. Heilbron

Georg Christoph Lichtenberg (1742–99) has left his name in a few obscure places in the rambling structure of physical science. You can find Lichtenberg's ring on the moon, near the Ocean of Storms; Lichtenberg figures, a precocious anticipation of xerography, in manuals of electrostatics; and Lichtenberg's alloy, of bismuth, lead, and tin, in handbooks of metallurgy. These accomplishments are perhaps insufficient to justify recalling Lichtenberg in this year of physics. What earns him this honour is his unusual pedagogy and his wit, two things that rarely occur together.

Lichtenberg became professor of mathematics and physics at the Georg-August University of Göttingen in 1775. Göttingen, the seat of learning of the electorate of Hanover, had close ties with Britain through George III, who put money into its library and sent his sons to study there. During visits to England in the 1770s, which he undertook at the invitation of English students he had tutored, Lichtenberg caught a bad case of Anglophilia. He became the world expert on the paintings and drawings of William Hogarth; George III made him a privy councillor in 1788; and the Royal Society elected him a fellow five years later. Despite a hunched back, the stature of a dwarf, and fragile health, Lichtenberg was a man of great charm. People as diverse as Alessandro Volta and Immanuel Kant admired him.

Lichtenberg gave lectures on experimental physics, which interested even the royal princes. As a teacher he employed a large variety of instruments and machines, and a textbook that he adapted from the *Anfangsgründe der Naturlehre* composed by his colleague Johann Polykarp Erxleben. Lichtenberg's version was arranged on unusual pedagogical principles. For one thing, it contained a little mathematics, which was not common in physics texts of the time. But the true novelty was that erroneous or superseded statements were retained, together with their corrections, in its successive updated editions.

In this way, Lichtenberg taught physics and modesty together. To err of course is human. In one of the aphorisms in the collection known as *The Waste Books*, Lichtenberg gave this truth a typical sardonic twist: "To err is human also insofar as animals seldom or never err, or at least only the



'Lichtenberg's figures' emerged from patterns formed by dust settling on a charged dielectric plate.

cleverest of them do so". Earlier this year the Academy of Sciences of Göttingen issued a very scholarly variorum edition of the *Anfangsgründe* as *Vorlesungen zur Naturlehre*. It will be a fine research instrument for historians. It is unlikely, however, to bring back teaching by disclosure of the teacher's errors.

Lichtenberg regarded the physical hypotheses or models of his time (a century after Newton!) as childish if not infantile. "In physics we have not yet reached puberty." He foresaw the adult argument that mathematical elegance and simplicity were indicators of truth, and he rejected it. "The lofty simplicity of nature all too often rests on the plain simplicity of the one who thinks he sees it."

He also foresaw, and approved, the strategy of building ever-larger instruments. He constructed a big electrophorus based on the design used by its inventor Volta, his great friend and a ladies' man ("ein Reibzeug für die Damen"), just to see what he could find. Having charged the dielectric plate in the usual way, by rubbing it with a docile cat, he saw the dust and shavings he had previously scraped off the plate fly to it and arrange themselves in intriguing patterns. The inventor of the Xerox technique, Chester Carlson, a patent lawyer as well as a physicist, recognized Lichtenberg's patterns as "the first electrostatic recording process".

Benjamin Franklin's lightning rods made no headway in the German states until Lichtenberg intervened in 1780.

Before then, churches had continued their ancient practice of ringing bells to break up thunderclouds, taking a high toll of bell ringers. In 1779, the year before Lichtenberg raised a lightning conductor constructed according to Franklin's instructions, a doctor in Hamburg had tried to construct one. But as he had neglected to ground his conductor, it was as dangerous as a damp rope in a wet belfry.

Lichtenberg not only made clear the functioning of the lightning rod, but also brought to light the "return stroke" — the current that flows through wet ground when a lightning flash annuls the charge induced on the Earth's surface. This current can extend some distance from the point struck. Consequently, as Lichtenberg advised in his droll way, when caught in a field in a thunderstorm, you must walk with your feet together to prevent the return stroke from passing up one leg and down the other.

Lichtenberg's aphorisms, jibes, digs and wisdom are available in modern editions in several languages. Today's English-speaking physicist may find much refreshment in R. J. Hollingdale's excellent translation of *The Waste Books*. Here are some examples. On modesty: "Physicians should say, not 'I cured him,' but 'he did not die at my hands.' In the same way one should not say in physics, 'I have explained the effect,' but 'I have assigned to it causes whose absurdity no one has as yet been able to demonstrate.'" On causality: "We have to believe that everything has a cause, as the spider spins its web to catch flies. But it does this before it knows that there are such things as flies." On anthropic hypotheses: "He marvelled at the fact that cats had two holes cut in their fur at exactly the spot where their eyes were." On quantum mechanics: "I have remarked very clearly that I am often of one opinion when I am lying down and of another when I am standing up." On theorists: "Not only did he not believe in ghosts, he wasn't even afraid of them." On academic blockheads: "*Non cogitant, ergo non sunt.*"

John L. Heilbron is at the Museum for the History of Science, Broad Street, Oxford, UK.

FURTHER READING

Lichtenberg, J. C. *The Waste Books* (transl. Hollingdale, R. J.) (New York Review of Books, New York, 1990).
Stern, J. P. *Lichtenberg: a doctrine of scattered occasions, reconstructed from his aphorisms and reflections* (Indiana Univ. Press, Bloomington, 1959).

UNIV. OF GÖTTINGEN

ESSAY

NEWS & VIEWS

D. BUTOW/CORBIS SABA



INTELLIGENCE

A gender bender

Steve Blinkhorn

The conclusion of a number-crunching exercise on various data sets is that male university students have significantly higher IQs than their female counterparts. But the methodology used is deeply flawed.

“Finding bad reasons for what we believe on instinct” is how the philosopher F. H. Bradley famously defined metaphysics¹. Meta-analysis — the deployment of statistical methods on a range of other people’s research results to draw stronger conclusions than any individual study can bear — is in jeopardy of coming into that kind of disrepute when it is treated in the manner of a paper just published in the *British Journal of Psychology*².

The paper is by Paul Irwing and Richard Lynn, and was widely trailed in the British press at the end of August. It claims to show an average IQ advantage of 4.6 points in favour of males among university students — about one-third of the standard deviation of IQ in the population as a whole, which by convention is set at 15. If this were true, it would really matter, in a way that resonates with the recent controversy at Harvard University concerning the representation of women in science³. It would also overturn a consensus of more than 50 years’ standing, that the only sex difference in IQ is a possible slightly greater variance among males.

The prospect of further controversy focused

on gender politics and the shortfalls of IQ tests does not seem remote. But the real discussion should concern whether this paper stands up to scrutiny. It does not.

The authors chose to investigate sex differences in scores on two versions of Raven’s matrices, the standard and the advanced. These are great survivors among intelligence tests, being still widely used in the same form around 60 years after they were developed. Their particular claims to fame arise from the fact that the content is entirely diagrammatic (and so, according to some, less culturally biased), and from the results of factor analyses that suggest the ability they tap is close to pure *g*, variously called general intelligence, fluid intelligence or the ability to learn.

Irwing and Lynn searched the literature for studies on university students that reported average scores by gender. They found 22, with sample sizes ranging from 30 to more than 9,000, publication dates from 1964 to 2004, and locations as various as Egypt, India, Mexico, Belgium, Australia and the United States — but curiously nothing from Britain where the tests were developed. For the most part,

the purposes of the studies identified were not to investigate sex differences.

The number-crunching phase of the meta-analysis — weighting differences by sample size — resulted in an estimated difference of 0.15 standard deviations in favour of males. There are technical reasons for supposing that even this is an overestimate, but the authors go on to find reasons for not accepting this result. First, they exclude the largest study of all, an explicit ‘norming’ exercise in Mexico that accounts for almost 45% of the data, on the grounds that it is an ‘outlier’, with a difference of only 0.07 of a standard deviation — males answering on average only 0.4 more items correctly out of 60.

Abandoning the principle of weighting results by sample size, Irwing and Lynn then take the median of the estimated differences (0.31), and multiply it by the general-population standard deviation of IQ (15.0) to get the estimate of a 4.6-point IQ advantage for males. Before going on to speculate on the neurology underlying the female inferiority that they claim to have demonstrated, they propose that this result is “strong corroboration” of other meta-analytical results they

have obtained for the general population.

It is nothing of the sort. The ten studies with estimated differences above the median cover a total of only 2,591 participants, whereas the ten studies with differences below the median account for 15,735 participants — the four largest differences come from samples of 111, 173, 124 and 300, the four smallest from samples of 844, 172, 9,048 and 1,316. Choosing to use the median is a flawed and suspect tactic.

But even going along with Irwing and Lynn's approach produces a different outcome if more plausible parameters are chosen. Whatever the standard deviation of IQ among university students, it is highly unlikely to be as large as 15, the conventional figure for the population as a whole. Raven's matrices, by design, tend not to produce normal (gaussian) distributions. An estimated standard deviation of 10 IQ points among students would ensure that there are not too many genuinely mentally defective undergraduates, even if some of them do have learning difficulties. This suggests a mean difference of up to around 1.5 IQ points — 10×0.15 (the figure weighted by sample size). Were a perfect study possible, with proper attention to sampling and motivational issues, I would expect even that to turn out to be an overestimate.

Raven's matrices have been widely used for decades, and it is likely that the 'file-drawer' effect is at work here — studies with no

significant or substantial results never find their way into a journal, with no reporting of separate statistics for each gender when there is no difference. My own file drawer turned out to contain an analysis of data from 752 applicants for places on one degree course during the 1970s, tested on the advanced matrices, which were designed for the top 20% of the population. This yielded an advantage of 0.07 standard deviations for females. The sample is larger than all but five of those found by Irwing and Lynn. Should it be included in a meta-analysis? Of course not: the data were collected for another purpose and, without attention to purpose and context, meta-analysis is just numerical manipulation for its own sake.

Where there are sex differences to be found, detailed study of the internal workings of the test tends to show why. That's not based on instinct, but on my professional experience in designing gender-fair tests. Meta-analysis is not a substitute for properly designed research, and sex differences in average IQ, if they exist, are too small to be interesting. ■

Steve Blinkhorn is at Psychometric Research and Development Limited, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK. e-mail: steve@prd.co.uk

1. Bradley, F. H. *Appearance and Reality: A Metaphysical Essay* (Swan Sonnenschein, London, 1893).
2. Irwing, P. & Lynn, R. *Br. J. Psychol.* **96**, 505–525 (2005).
3. Singer, E. *Nature* **434**, 424 (2005).

ASTRONOMY

Light on a dark place

Christopher Reynolds

The sharpest images ever taken of matter around the probable black hole at the centre of our Galaxy bring us within grasp of a crucial test of general relativity — a picture of the black hole's 'point of no return'.

Ever since its discovery in 1974, a strange source of radio waves in the constellation of Sagittarius has been suspected of flagging the presence of a massive black hole at the centre of our Galaxy¹. On page 62 of this issue, Shen *et al.*² report the highest-resolution images yet of this source, Sgr A*. These observations provide strong evidence that Sgr A* is indeed a black hole, and afford a glimpse of the behaviour of the matter that is about to flow into it. They are also a further step towards attaining an image of the shadow around the edge of a black hole, a powerful and classic test of the general theory of relativity.

Black holes are perhaps the most exotic objects to impinge on the cosmic consciousness. They are formed when matter such as that from a dying massive star collapses in calamitously under its own gravity, forming a region of space in which the gravitational field is so strong that it swallows all matter and radiation

that come near it. Delineating this region is the event horizon: the point of no return, beyond which no matter or light can ever escape.

Sgr A* is certainly an exotic object. Through observations at infrared frequencies of the bright stars speeding around it, astronomers have confirmed that it is four million times more massive than our Sun and confined to a region of space at the exact centre of the Galaxy^{3,4} that is no bigger than the region enclosed by the orbit of Pluto. Such a high concentration of mass puts tight constraints on the possible nature of the object. A cluster of several million neutron stars, themselves collapsed dead stars, could be as heavy as that, but could only survive in such a compact form for 20,000 years or so — a blink of an eye in astronomical terms — before either collapsing further (to a black hole) or evaporating away into space. It is unlikely that we are observing the galactic centre just when such a bizarre

neutron-star cluster happens to exist. There is only one other possibility, however, if standard physics — the standard model of particle physics, coupled with the general theory of relativity — is to hold. That is that Sgr A* harbours a supermassive black hole.

Shen *et al.*¹ use a technique known as Very Long Baseline Interferometry (VLBI), which correlates information from radio antennae at separate, remote locations, thereby increasing the spatial resolution for images of far-off objects. The authors used the Very Long Baseline Array, a system of ten radio telescopes scattered across the United States with a maximum separation of some 8,000 kilometres. They built up a picture of the radio emission at a wavelength of 3.5 centimetres from gas in a region just 8 light minutes across, centred on the putative black hole. Even with the most conservative assumptions, the authors find lower limits on the concentration of mass that are a factor of 100,000 higher than those derived from the motions of the stars surrounding it. This would reduce the lifetime of any neutron-star cluster there to a mere 100 years, a result that must dispel any lingering notions that the source at Sgr A* is a compact cluster of known objects.

But we should guard against complacency: nature might have some surprises in store. Could it be that standard physics is inadequate and that, other than a black hole, there are stable objects that have the compact, huge mass of Sgr A*? What is needed is a more discerning test than simply detecting something massive and compact: we need to find the event horizon, the defining property of a black hole.

As physical phenomena go, event horizons are tricky to observe. In fact, the existence of an event horizon is almost invariably argued for based on the absence of some observational signature^{5,6}. High-resolution imaging, however, does provide a compelling way to search for an event horizon. If a black hole is surrounded by an almost spherical distribution of radiating matter, as in the case of Sgr A*, a sufficiently high-resolution image should reveal a shadow around it (Fig. 1). This dark circle is caused by radiation from sources behind the black hole that are being swallowed by the event horizon⁷. Surrounding this shadow would be a bright ring — the result of the strong deflection by the black hole's gravitational field of those light rays that do scrape past it.

The predicted diameter of the event horizon's shadow for Sgr A* is just 30 microarcseconds⁷, or 120 millionth of a degree. This would be the apparent size of a tennis ball on the Moon when viewed from Earth, and is about a factor of four smaller than the scales probed by current VLBI experiments. Two main directions of research should eventually allow us to see an image of the event horizon — if indeed there is one.

First, the resolving power of VLBI can be improved by reducing the wavelength at which



Figure 1 | Hungry mouth. The event horizon of a black hole (the bright area at the centre of the image) completely absorbs emissions from matter behind the black hole when viewed from Earth. The result is a darker circle that could be seen in a sufficiently high-resolution image. In the case shown here — a theoretical calculation of the event-horizon shadow of the source Sgr A*, which is the subject of Shen and colleagues' study¹ — the black hole is assumed to be rotating rapidly. The tendency of photons to be flung around the black hole in the direction of its rotation brings about an event-horizon shadow that is off-centre, to the right in the image. A brighter ring around the shadow is formed by light rays that are strongly deflected by the gravitational pull of the black hole without being absorbed by it. (Figure courtesy of Eric Agol, Univ. Washington.)

the observations are made, and thus the effects of diffraction that fundamentally limit the resolving power of any telescope. Observations at shorter wavelengths are required in any case: the emitting gas around the event horizon in Sgr A* is opaque to radiation of wavelengths longer than around 1 mm. VLBI observations at submillimetre wavelengths capable of detecting the shadow will become feasible within the next five to ten years. The launch of spacecraft such as NASA's iARISE will also extend the effective size of the telescopic system beyond the limits imposed by our planet.

Second, and on a longer timescale of perhaps

30 years, NASA plans to develop an X-ray interferometer that will allow extremely high-resolution imaging of matter in Sgr A* that emits at even shorter X-ray wavelengths. It will also image accreting massive black holes in nearby galaxies. Further data should emerge from future generations of large X-ray telescopes such as NASA's Constellation-X and the European Space Agency's XEUS missions, and from the ground-breaking work in gravitational-wave astronomy currently exemplified by the LIGO, VIRGO and LISA interferometers. Together with these experiments, high-resolution imaging will herald a new era in probing the structure and

properties of some of the most enigmatic objects in the Universe.

Christopher Reynolds is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

e-mail: chris@astro.umd.edu

1. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
2. Shen, Z.-Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P. & Zhao, J.-H. *Nature* **438**, 62–64 (2005).
3. Ghez, A. M. et al. *Astrophys. J.* **620**, 744–757 (2005).
4. Schoedel, R. et al. *Nature* **419**, 694–696 (2002).
5. Garcia, M. et al. *Astrophys. J.* **553**, L47–L50 (2001).
6. Narayan, R. & Heyl, J. S. *Astrophys. J.* **574**, L139–L142 (2002).
7. Falcke, H., Melia, F. & Agol, E., *Astrophys. J.* **528**, L13–L16 (2000).

MICROBIOLOGY

Algae and the vitamin mosaic

Robert A. Andersen

The requirements for vitamin B₁₂ vary among algal species in a seemingly inexplicable pattern. A study that exploits genomic data now provides enlightenment — and evidence of symbioses with bacteria.

Animals require vitamin B₁₂ in their diets but plants don't. Algal requirements, on the other hand, present a complicated picture. Many algae must take up vitamin B₁₂ (like animals) whereas many others do not (like plants)¹, and there is no evolutionary pattern to these requirements. Even different isolates of the same species can have different demands².

On page 90 of this issue³, Croft *et al.* provide an explanation for this baffling mosaic of algal vitamin-B₁₂ requirements. And they go a step further. For algae that require an external source of vitamin B₁₂, they provide convincing evidence of an algal–bacterial symbiosis in which algal carbon-rich exudates are exchanged for the vitamin produced by the bacteria. For these algae, vitamin B₁₂, they find, is an essential co-factor for an enzyme on a pathway that produces the amino acid methionine.

When algal species were first cultured more

than 100 years ago, they were grown on a defined plant culture medium that lacked vitamins⁴. But some algae could not be grown in this way, and soil extract was added to the medium for these 'unculturable' organisms⁵. The soil provided, among other nutrients, an unrecognized source of vitamins, making it possible to culture most species of algae. In 1948, liver extract was shown to improve the growth of *Euglena*⁶ and evidence for a vitamin-B₁₂ requirement in algae came one year later⁷. Shortly thereafter, *Euglena* and *Potriochromonas* became important bioassay organisms for detecting vitamin-B₁₂ deficiencies in human blood and urine samples¹. By 1980, vitamin-B₁₂ requirements were known for approximately 400 algal strains⁸, but after this, studies on the topic almost ceased.

Croft *et al.*³ have picked up the baton. They started by re-examining the vitamin-B₁₂

requirements of some algal strains. Their results were identical to previous reports. Puzzled by the lack of an evolutionary pattern for vitamin-B₁₂ requirements among algae, they searched for genes encoding vitamin-B₁₂-dependent enzymes in the genomes of the green alga *Chlamydomonas*, the red alga *Cyanidioschizon* and the brown diatom *Thalassiosira*.

They found that *Chlamydomonas* and *Cyanidioschizon* have both a vitamin-B₁₂-dependent methionine synthase gene (*metH*) and a vitamin-B₁₂-independent methionine synthase gene (*metE*). Both organisms can grow without an external vitamin source. Croft *et al.* further showed that *Chlamydomonas* preferentially used the *metH* gene when vitamin B₁₂ was available, but used the *metE* gene when vitamin B₁₂ was absent. Conversely, they found only the vitamin-B₁₂-dependent methionine synthase gene (*metH*) in the *Thalassiosira* genome, an alga with an absolute requirement for this vitamin⁹.

Presumably, the ancestors of algae had both genes, and numerous independent losses of either the *metE* or *metH* gene subsequently occurred over evolutionary time. Remarkably, descendants seem to have continued to arise almost exclusively from ancestors with both genes — which explains why we do not find lineages (as we do with plants and animals)

with either one gene type or the other. The scattered losses produced the evolutionary mosaic at the level of species or even strain.

Croft *et al.*³ also provide evidence that a bacterium, *Halomonas*, upregulates the biosynthesis of vitamin B₁₂ when in the presence of algal exudates. Many algae are members of a loose taxonomic grouping known as the protists. These are often unicellular and include such organisms as *Amoeba* and *Paramecium*. The phenomenon of endosymbiosis, in which one organism (such as a bacterium) takes up residence inside another, to mutual benefit, has been thoroughly studied in protists, especially with regard to the origins of the chloroplast and mitochondrion. But possible ectosymbioses — literally, more superficial relationships — involving bacteria and algae have received less attention.

An earlier investigation did indeed show that *Thalassiosira* and other marine diatoms, all of which require vitamin B₁₂, could be grown without the vitamin when bacterial cultures were added to the diatom cultures¹⁰. Such studies hinted at the existence of a symbiotic relationship. But unlike Croft *et al.*, the authors of this study did not identify the bacteria involved or the specific genes (or enzymes) concerned, and they did not demonstrate upregulation of a bacterial gene in response to a chemical signal from the algae.

Croft and colleagues' approach³ could profitably be adopted more broadly, because protists have a much wider variety of basic biochemical pathways than do either animals or plants. We can hope that the enzymatic pathways leading to other amino acids, sugars, lipids and so forth — which have long been known to be diverse in protists and to show similar evolutionary mosaic patterns¹¹ — will likewise be examined using the genome data now available. It is likely

that additional symbiotic vitamin-B₁₂-producing bacteria will be identified, and that other vitamins are produced by symbiotic bacteria.

But non-symbiotic bacterial sources of vitamins may be equally or more important. For example, concentrations of vitamin B₁₂ in the oceans vary with season, and there is strong circumstantial evidence that this vitamin is produced on the ocean floor at depths where darkness makes it unlikely that an algal-bacterial symbiosis can exist¹².

Clearly, the paper by Croft *et al.* doesn't answer all questions. But it greatly advances our understanding of why the vitamin-B₁₂ requirements are so sporadic among the algae, and also points to an enticing variety of research opportunities.

Robert A. Andersen is at the Provasoli-Guillard National Center for Culture of Marine Phytoplankton, Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine 04575, USA.

e-mail: randers@bigelow.org

1. Provasoli, L. & Carlucci, A. F. in *Algal Physiology and Biochemistry* (ed. Stewart, W. D. P.) 741–787 (Blackwell, Oxford, 1974).
2. Lewin, J. C. & Lewin, R. A. *Can. J. Microbiol.* **6**, 127–134 (1960).
3. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. *Nature* **438**, 90–93 (2005).
4. Faminzain, A. *Bull. Acad. Sci. St. Petersburg* **17**, 31–70 (1871).
5. Pringsheim, E. G. *Beitr. Biol. Pfl.* **11**, 305–334 (1912).
6. Provasoli, L., Hutner, S. H. & Schatz, A. *Proc. Soc. Exp. Biol. Med.* **69**, 279–282 (1948).
7. Hutner, S. H. *et al. Soc. Exp. Biol. Med. Proc.* **70**, 118–120 (1949).
8. Swift, D. in *The Physiological Ecology of Phytoplankton* (ed. Morris, I.) 329–368 (Univ. California Press, Berkeley, 1980).
9. Guillard, R. R. L. & Ryther, J. H. *Can. J. Microbiol.* **8**, 229–239 (1962).
10. Haines, K. C. & Guillard, R. R. L. *J. Phycol.* **10**, 245–252 (1974).
11. Ragan, M. A. & Chapman, D. J. *Biochemical Phylogeny of the Protists* (Academic, New York, 1978).
12. Menzel, D. & Spaeth, J. P. *Limnol. Oceanogr.* **7**, 151–154 (1962).



50 YEARS AGO

"Use and abuse of English in science" — Another problem which is causing increasing concern — to printers as well as to editors — is the frequent and indiscriminate use of abbreviations in the form of a single capital letter, or a group of capitals, to represent the name of a substance, or perhaps even an adjective or adverb. The printer is concerned because a page of text sprinkled with capital letters is not pleasing in appearance; and like other craftsmen, he feels that his efforts are being frustrated... The use of abbreviations, especially initial letters, is now becoming so fashionable among scientists that one suspects authors sometimes go out of their way to use them. ... this fashion may, if not checked, defeat its own ends and produce a veritable 'Tower of Babel'. Indeed the time does not seem far away when high-school pupils will have to learn a new table of symbols apart from those atomic.

From *Nature* 5 November 1955.

100 YEARS AGO

A return has been published, we learn from the *Pioneer Mail*, regarding the measures adopted for the extermination of wild animals and venomous snakes during the year 1904. The total mortality among human beings reported to have been caused by wild animals was 2157, against 2749 in 1903. The most notable decrease occurred in Madras and the United Provinces, namely, from 438 and 404 in 1903 to 237 and 193 in 1904 respectively... The mortality from snake-bite rose from 21,827 to 21,880. It is reported that in the Seoul district of the Central Provinces anti-venin was used with success in two cases, and the question of introducing more generally the treatment of snake-bite by potassium permanganate is under the consideration of the local Government. The total number of snakes killed was 65,378.

From *Nature* 2 November 1905.

GLOBAL CHANGE

Sea level and volcanoes

Anny Cazenave

Large volcanic eruptions cool the world ocean. In doing so, they temporarily reduce the increase in ocean heat content and the rise in sea level attributed to warming caused by greenhouse-gas emissions.

Global warming is producing a rise in sea level. Observations from tide gauges and satellite altimetry indicate that sea level has been rising by 1.8 millimetres per year since 1950 (ref. 1) and about 3 millimetres per year during the 1990s (ref. 2). The two causes are a thermal expansion of sea water in response to ocean warming, and the input of extra water from the melting of glaciers and ice sheets on land³. But against the background of this overall increase, global mean sea level displays interannual to decadal oscillations of the order

of several millimetres. These oscillations have received scant attention to date. Some of them result from changes in ocean heat content associated with internal perturbations of the ocean-atmosphere system, such as the El Niño–Southern Oscillation and Pacific Decadal Oscillation⁴. But other processes, probably related to the 'forcing' effects of natural climate variation, also play a role.

On page 74 of this issue, Church *et al.*⁵ use climate simulations to reveal the effects of volcanic eruptions on sea level between 1890 and

50 & 100 YEARS AGO



Figure 1 | Mount Pinatubo erupts. The cooling effect of the sulphate aerosols produced by this and other eruptions helps to account for oscillations in the sea-level record⁵.

2000. The simulations take account of anthropogenic forcing that stemmed from greenhouse gases, aerosols and ozone, and the natural climate forcing resulting from changes in volcanic activity and the input of solar radiation. These forcings affect the oceans through warming (or cooling) the upper ocean, leading to an increase (decrease) of ocean heat content and hence an increase (decrease) in sea level through ocean thermal expansion (contraction). Church *et al.* find that, in the first few months after large volcanic eruptions, there is a fall of several millimetres in global mean sea level. This is followed by a slow increase, lasting for a decade or more, towards the pre-eruption state.

The link here is that large volcanic eruptions inject particles and gases into the atmosphere, in particular sulphur gases that are converted into sulphate aerosols in the stratosphere, the layer of Earth's atmosphere immediately above the lowest layer, the troposphere. Their dominant effect is to increase the fraction of incident radiation reflected by the planet, and hence reduce the amount of solar energy reaching Earth's surface. The corresponding net surface-air cooling and its consequences on weather have been much studied⁶, unlike the impact on ocean heat content and sea level.

Church *et al.*⁵ show that, because of the reduction of the net solar flux at the ocean surface, volcanic eruptions induce an immediate cooling of the surface layers, and so a decrease in heat content and sea level. The predicted abrupt fall in ocean heat content is well correlated with observations⁷. Although surface-air temperature recovers within a few years, the cooling effects on the ocean persist for at least a decade. This is because of the large heat capacity of the oceans compared with that of the atmosphere and the slow redistribution of heat by the ocean circulation^{5,8}.

Observational analyses of historical ocean temperatures show a net warming of the oceans since 1950, contributing about 85% of the total increase in heat content of the whole Earth system⁷ and consistent with the current

imbalance between the energy absorbed and emitted by the planet⁹. Model studies suggest that most of this ocean warming results from human activities and the associated increase in levels of greenhouse gases¹⁰. However, during the past several decades, large volcanic eruptions — Mt Agung, Indonesia (1963), El Chichon, Mexico (1982) and Mt Pinatubo, Philippines (1991) — have temporarily reduced the anthropogenic ocean warming. Because the effects of subsurface ocean cooling can persist for one to several decades, they have masked at least part of the accelerating increase in sea level.

The simulations⁵ indicate that the eruption of Mt Pinatubo in 1991 (Fig. 1) produced a drop of some 6 mm in sea level within about a year, which was followed by a slow rise of about 0.5 mm yr⁻¹ over the next decade or more. Thus, about 0.5 mm yr⁻¹ of the steeper sea-level rise resulting from thermal expansion, estimated from ocean temperature data over the past decade (about 1.5 mm yr⁻¹, compared with the 0.4 mm yr⁻¹ mean rate of the

past 50 years)¹¹, may reflect a recovery from the Mt Pinatubo eruption. This in turn may explain part of the higher rate of sea-level rise observed by satellite altimetry since early 1993 (3 mm yr⁻¹, compared with 1.8 mm yr⁻¹ recorded by historical tide gauges since 1950; ref. 1). Although it remains unclear whether sea-level rise over the past decade indicates an accelerating trend, the study by Church *et al.*⁵ suggests that part of it (although not all) can be explained by natural variability on interannual to decadal timescales.

Climate-model predictions indicate that sea level will continue to rise in the coming decades, and even centuries, owing to thermal expansion of the ocean in response to anthropogenic warming³. But Church *et al.*⁵ clearly demonstrate that large volcanic eruptions can (partially and temporarily) mask that effect. Their study is one step towards a better understanding of the sea-level record — which is essential if we are to improve projections of sea-level rise, and prepare for the impact of that rise on vulnerable coastal regions and island nations.

Anny Cazenave is at the Laboratoire d'Etudes en Géophysique et Océanographie Spatiales, LEGOS-CNRS, 18 avenue Edouard Belin, 31401 Toulouse, Cedex 9, France.

e-mail: anny.cazenave@cnes.fr

1. Church, J. A., White, N. J., Coleman, R., Lambeck, K. & Mitrovica, J. X. *J. Clim.* **17**, 2609–2625 (2004).
2. Leuliette, E. W., Nerem, R. S. & Mitchum, G. T. *Mar. Geodesy* **27**, 79–94 (2004).
3. Church, J. A. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 639–694 (Cambridge Univ. Press, 2001).
4. Lombard, A., Cazenave, A., Le Traon, P. Y. & Ishii, M. *Glob. Planet. Change* **47**, 1–16 (2005).
5. Church, J. A., White, N. J. & Arblaster, J. M. *Nature* **438**, 74–77 (2005).
6. Robock, A. *Rev. Geophys.* **38**, 191–219 (2000).
7. Levitus, S., Antonov, J. I. & Boyer, T. P. *Geophys. Res. Lett.* **32**, L02604; doi:10.1029/2004GL021592 (2005).
8. Delworth, T. L., Ramaswamy, V. & Stenchikov, G. L. *Geophys. Res. Lett.* (in the press).
9. Hansen, J. *et al.* *Science* **308**, 1431–1435 (2005).
10. Barnett, T. P. *et al.* *Science* **309**, 248–287 (2005).
11. Antonov, J. I., Levitus, S. & Boyer, T. P. *Geophys. Res. Lett.* **32**, L12602; doi:10.1029/2005GL023112 (2005).

STRUCTURAL BIOLOGY

Proteins flex to function

Yuanpeng J. Huang and Gaetano T. Montelione

Static pictures of protein structures are so prevalent that it is easy to forget they are dynamic molecular machines. Characterizing their intrinsic motions may be necessary to understand how they work.

Enzymes are proteins that catalyse chemical processes by lowering the energy barriers between substrate and product, thus increasing the rate of reaction. They function by stabilizing transition states, the structural intermediates formed in the rate-limiting steps of chemical reactions. In some cases, the catalysis process involves the enzyme interconvert-

ing between alternative versions of its structure, which can be followed using nuclear magnetic resonance (NMR) spectroscopy^{1–3}. On page 117 of this issue, Eisenmesser *et al.*⁴ demonstrate that the slow structural motions between sub-states of an enzyme that take place during catalysis also occur in the free enzyme in the absence of its substrate. This

implies that enzymes have a set of pre-existing dynamics, or 'intrinsic plasticity', that can determine their function.

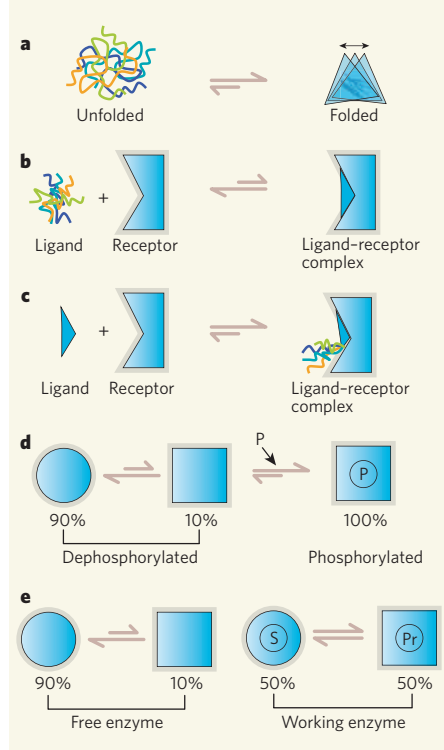
The conventional view of protein structure is a static arrangement of atoms in space. For example, structures derived from X-ray crystallographic studies are usually reported as the single set of atomic coordinates that best fit the experimental diffraction data. The structural biologists who carry out such studies recognize that, even in the crystalline state, there are thermal vibrations and collective dynamics within these structures. However, the idea that protein structure is rigid and immobile is propagated by the convention of presenting a single atomic model for the protein and by the limitations of print-based publication — it is difficult to convey dynamic structural information on the pages of a journal.

During the past 30 years, spectroscopic studies of proteins, particularly vibrational and NMR spectroscopy, have provided a wealth of data about internal dynamics and conformational sub-states^{5–8}. Modern NMR experiments can monitor molecular motions quantitatively on nanosecond or on micro- to millisecond timescales^{7,8}. Such studies reveal that proteins are highly flexible, not only through movements of surface loops and side chains, but also by collective motion of the core structure.

Intrinsic dynamics can be linked to a protein's function in several ways (Fig. 1a–d). Changes in conformational dynamics between folded and unfolded states contribute a significant entropy component to the energetics of protein stability. Similarly, entropic effects due to changes in internal dynamics associated with molecular recognition processes can have a profound impact on binding affinities. Chemical modification of a protein, for instance by the addition of a phosphate group, can also stabilize one of several interconverting structures from a dynamic equilibrium of protein structures⁹. In enzymes, such transitions between structural sub-states can correspond to the rate-limiting step of catalysis^{1–4}.

Eisenmesser *et al.*⁴ have identified another class of dynamics to be taken into account when considering protein structure (Fig. 1e). They used NMR to follow the structural changes of the enzyme cyclophilin A during catalysis and in its substrate-free state. During catalysis, as the enzyme moves back and forth between sub-states, a subset of amino-acid residues experiences several local environments that interconvert with an exchange rate (the sum of the rates for the forward and reverse reactions) of $2,730 \pm 220 \text{ s}^{-1}$. This exchange rate is very close to that associated with substrate catalysis — that is, $2,500 \pm 500 \text{ s}^{-1}$ — indicating that these slow conformational changes of the enzyme coincide with substrate turnover¹.

Notably, there are similar characteristic dynamics in the same regions of the cyclophilin A structure even in the absence of



substrate. The free cyclophilin A had a conformational interconversion exchange rate of $1,140 \pm 200 \text{ s}^{-1}$. Once the difference in populations of the two interconverting states is taken into account, this rate is similar to that observed for the enzyme during catalysis. NMR chemical-shift data characterizing these interconverting states are also similar in the free and working enzyme, supporting the view that this is in fact the same dynamic equilibrium.

The collective motions are not just around the enzyme's active site, but involve an extensive region of the core protein structure. Eisenmesser *et al.*⁴ created several single-site mutants of cyclophilin A to explore this network of internal motions. The mutations do not affect either the size or the exchange rate of the dynamic network, but do cause the conformational equilibrium to shift among the sub-states — so that the sub-state populations in Fig. 1e might change to 80% and 20%, for example. The authors suggest that the catalytic power of the enzyme might be limited by the rates of these conformational rearrangements. A loop region of the free cyclophilin A also exhibits evidence for conformational exchange with a faster rate of $2,260 \pm 200 \text{ s}^{-1}$, demonstrating the complexity of the internal dynamics.

The similarity of the movements of free and working cyclophilin A implies that the protein motions associated with catalysis are an intrinsic property of the enzyme itself, dictated by the amino-acid sequence and selected for by molecular evolution. Indeed, the protein structure seems poised to undergo the catalytic process, because the numerous structural sub-states involved in this process are present in the free enzyme, although with

Figure 1 | The purpose of plasticity. The intrinsic dynamics of a protein are fundamental to its structure and contribute to how it works. The double arrows indicate interconverting states, with the larger arrows showing the direction that is most favoured energetically. Percentages give rough populations for each state. **a**, Changes in protein dynamics upon folding, and residual intrinsic dynamics in the 'folded' state of the protein, can modulate protein stability. **b, c**, Changes in intrinsic protein dynamics associated with ligand binding can modulate binding affinities. **b**, Some proteins increase their internal flexibility (or even unfold) upon dissociation of a complex, to weaken binding affinities. **c**, Other proteins retain (or create) structural dynamics upon complex formation, reducing the loss of entropy that is associated with molecules binding together¹⁰, further stabilizing the complex. **d**, Intrinsic dynamics of a protein can include sub-states that are stabilized by chemical modification, such as phosphorylation (P). **e**, Eisenmesser *et al.*⁴ show that the enzyme cyclophilin A goes through the same motions in the absence of its substrate (S) as it does when it is carrying out catalysis to form a product (Pr). However, the populations of the two sub-states are different in the free versus the working enzyme — making formation of the product more favourable.

altered populations compared with the enzyme carrying out catalysis. Such considerations could be critical for enzyme design and engineering. But will the same be true for other enzymes? The rate-limiting step for many enzymes will involve different kinds of chemistry from that applied by cyclophilin A, such as bond breakage or diffusion-limited events. So whether cyclophilin A is typical in having an intrinsic plasticity that is fundamental to its function, or is a special case, remains to be seen. It is increasingly evident that static protein structure is only part of the story of protein function; the other half tells of mobility, and many chapters are left to be written on how the intrinsic dynamics of a protein's structure can provide an underlying basis for its biological functions.

Yuanpeng J. Huang and Gaetano T. Montelione are at the Center for Advanced Biotechnology and Medicine, Rutgers University and Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA.
e-mail: guy@cabm.rutgers.edu

1. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. *Science* **295**, 1520–1523 (2002).
2. Rozovsky, S., Jogi, G., Tong, L. & McDermott, A. E. *J. Mol. Biol.* **310**, 271–280 (2001).
3. McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J. & Wright, P. E. *Proc. Natl Acad. Sci. USA* **102**, 5032–5037 (2005).
4. Eisenmesser, E. Z. *et al.* *Nature* **438**, 117–121 (2005).
5. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. *Biochemistry* **14**, 5355–5373 (1975).
6. Wagner, G. & Wuthrich, K. *Nature* **275**, 247–248 (1978).
7. Palmer, A. G. *Chem. Rev.* **104**, 3623–3640 (2004).
8. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. *Nature Struct. Biol.* **8**, 932–935 (2001).
9. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. *Science* **291**, 2429–2433 (2001).
10. Lee, A. L., Kinnear, S. A. & Wand, A. J. *Nature Struct. Biol.* **7**, 72–77 (2000).

COSMOLOGY

The infrared dawn of starlight

Richard S. Ellis

The modest-sized but successful Spitzer Space Telescope has detected fluctuations in cosmic light at infrared frequencies. Is this the signature of the first population of stars that formed in the Universe?

On page 45 of this issue, Kashlinsky *et al.*¹ present observations that reveal clustering in the distribution of cosmic infrared light over and above that expected from the combined effect of known galaxies. This excess signal could conceivably be light from stars that switched on when the Universe was just a tiny fraction of its present age. The authors do not detect any individual sources, nor can they pinpoint precisely when in cosmic history these signals were produced. Nevertheless, their result is likely to provoke much discussion among cosmologists.

Searching for cosmic sources that ignited when the Universe was very young is at the frontier of our current observational capabilities. Deep surveys with space-based telescopes and the larger ground-based telescopes allow us to observe very faint sources that are thought to have originated in early cosmic times. Much progress has been made in tracing how early stellar systems changed and grew to become 'respectable' galaxies such as our own Milky Way. But the results of numerical simulations² suggest that the very first stars may have been quite different from those that came later. They probably contained only atoms of light nuclei produced in the Big Bang — so no heavy elements such as carbon and oxygen. They are also likely to have been very massive, over a hundred times more massive than the Sun. When these heavyweights switched on, they would have burnt up their hydrogen in only a few million years, shining intensely and briefly in a brilliant dawn of starlight.

Detecting and studying these first stars will

be a specific task for the 6.5-metre-aperture James Webb Space Telescope — successor to the 2.5-metre Hubble Space Telescope — that is projected to enter orbit in 2013. Kashlinsky and colleagues' observations¹ come from a more modest-sized instrument, the Infrared Array Camera (IRAC) on board NASA's Spitzer Space Telescope. The technique they use — that of analysing ripples in the background sky — is well established in cosmology: observations of ripples at microwave frequencies produced soon after the Big Bang (the so-called cosmic microwave background) gave us precise measurements of the constituents and geometry of the Universe. IRAC detects light at infrared wavelengths between 3.6 and 8 micrometres, and is sensitive to the light of young stars that has been redshifted (moved to longer wavelengths) by a factor of between 20 and 50 by the expansion of the cosmos since its emission. This light would have originated at a time when the Universe was only 100 million years old — an infant in terms of its stately 13.7 billion years today.

Spitzer, with its 85-centimetre aperture, does not have the angular resolution to resolve individual sources at these enormous distances. But it might well be capable of detecting distant stars' combined output as structure in the light produced behind that of foreground sources. And here lies the observational challenge: Spitzer's IRAC detectors are also sensitive to diffuse light that is produced within the Solar System, in clouds of cool gas in our Galaxy and, crucially, from galaxies — those close by, as well as those a long way back in cosmic time. Even a minor blunder in

removing these foreground signals might lead to a spurious result.

Although several groups have tentatively claimed to have seen fluctuations from early starlight on the basis of previous infrared space missions^{3,4}, the increased depth of the IRAC images means that Kashlinsky *et al.*¹ could reduce the confusing signals from foreground galaxies to fainter upper limits (Fig. 1). They present several tests to demonstrate the reliability of their detected signal and the origin they claim for it. The pattern of structure they see, for example, is statistically similar at four different infrared wavelengths, and is consistent across fields whose sky coordinates are very different (so an origin in the Milky Way is unlikely). The signal does not change between two measurements taken six months apart, as would be the case if its origin were 'zodiacal' light reflected by local interplanetary dust in the Solar System.

The authors use various methods to remove the signals coming from faint galaxies, none of which significantly alters their result. Remarkably, however, the total contribution of foreground galaxies, estimated by extrapolating to include those too faint to be resolved in the deepest IRAC images⁵, is small compared with the residual signal, which the authors ascribe to primordial stars. Kashlinsky *et al.* argue that this is further proof of their signal's origin, because only a very large error in the amount of foreground galaxy clustering would make the residual signal disappear. Because the clustering observed in the background will be diluted if the first stars switch on gradually over a range in cosmic times, the prominent signal might conceivably arise because it was produced in a very short time interval. But the imbalance still seems a surprising outcome, and a number of untested assumptions involved in allowing for unobserved galaxies could represent a weakness in the analysis.

Kashlinsky and colleagues, perhaps wisely, do not interpret their signal in much detail. Theorists have predicted the level of fluctuations that might be seen in such experiments⁶; but their calculations, too, depend on many imponderables. The genuine excitement in this work lies in the practicality of detecting the stellar radiation from hitherto uncharted distances corresponding to a time when the Universe was barely 100 million years old. Not bad for an 85-centimetre telescope!

Richard S. Ellis is at the California Institute of Technology, MS 105-24, Pasadena, California 91125, USA.

e-mail: rse@astro.caltech.edu

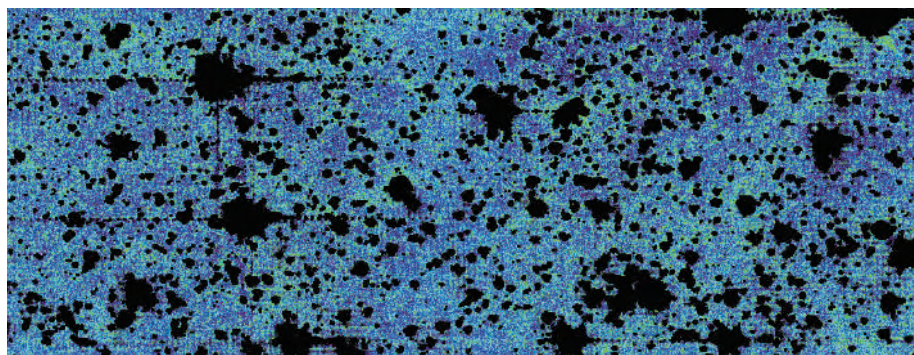


Figure 1 | Sign of the times? A deep infrared exposure taken with the Spitzer Space Telescope. When contributions from the known stars and galaxies are subtracted (black areas), the fluctuations in the background emission show a residual signal over and above that expected from the past history of galaxies. This may arise from the first massive stars that ignited when the Universe was only 100 million years old. (Courtesy of A. Kashlinsky *et al.*¹)

1. Kashlinsky, A., Arendt, R. G., Mather, J. & Mosely, S. H. *Nature* **438**, 45–50 (2005).
2. Abel, T. *et al.* *Science* **295**, 93–98 (2002).
3. Wright, E. L. & Reese, E. D. *Astrophys. J.* **545**, 43–55 (2000).
4. Matsumoto, T. *et al.* *Astrophys. J.* **626**, 31–45 (2005).
5. Fazio, G. *et al.* *Astrophys. J. Suppl. Ser.* **154**, 39–43 (2004).
6. Santos, M., Bromm, V. & Kamionkowski, M. *Mon. Not. R. Astron. Soc.* **336**, 1082–1092 (2002).

CHEMICAL BIOLOGY

Bring them back alive

Michael Yarus

A deep search has turned up an RNA that can carry out the chemically complex 'aldol' reaction involved in sugar metabolism. Could this be similar to an ancestral catalyst that existed billions of years ago?

The title of this piece evokes the capture of exotic beasts for zoos, and hints at the reanimation of the extinct. Both goals apply to the work of Fusz *et al.*¹, who, as they report in *Chemistry and Biology*, went hunting for a sort of molecule that might have existed about 4 billion years ago. According to the 'RNA world' hypothesis, life at this time consisted of macro-molecular or cellular assemblies, with RNA molecules rather than proteins catalysing chemical reactions. In support of this, a number of RNA enzymes (ribozymes) have been created, or captured from the wild, that can carry out the sorts of chemical reaction — the formation of peptide bonds, for example — that would have been necessary for such life.

Now Fusz *et al.* identify a ribozyme that can catalyse the particularly tricky aldol reaction. This involves carbon–carbon bond formation which would have been required at several significant points in the evolution of life, including, possibly, when ribose was first produced. Fusz and colleagues' success at finding a ribozyme that could catalyse the aldol reaction (Fig. 1) makes it more likely that the 'organisms' in the RNA world actually used aldol reactions².

The authors used selection–amplification, or SELEX³, to track down their prey. This technique sifts through a quadrillion or so random sequences of RNA to find a few (or perhaps just one) that can carry out a specified reaction. This may seem like looking for a needle in a haystack, but SELEX is extremely powerful: it has turned up more RNA catalysts than have been found by harvesting from living

material⁴. Aldol chemistry occurs in a crucial step in sugar metabolism in all cells, and is central to glycolysis, the Calvin cycle, and other energy-yielding pathways. As only the second reaction forming a carbon–carbon bond found to be carried out by RNA^{5,6}, aldol reactions are also a welcome addition to RNA catalytic diversity.

In nature, novel activities would probably arise from the adaptation of existing molecules that already had some of the required properties. By contrast, SELEX usually involves a single search of a set of randomized sequences. The number of sequences expands exponentially with the size of the molecule found — ten times for each 1.6 nucleotides added⁷. It is much easier for a novel activity to evolve in nature, through the selection of subtle enhancements of an existing molecule, than for us to happen upon an entirely new activity all at once by finding a sequence of 15–20 nucleotides. Thus, when single-step selection succeeds, it strongly implies that the active site could also have been found by ribocytes, the supposed primordial RNA 'cells', using a more effective evolutionary search.

Fusz *et al.* ensured that the desired aldol catalyst would announce itself by coaxing the aldol reaction to link a tag to the active RNA. The tag then aided the physical separation of the RNA from the substrate so that the RNA could be studied in isolation. The result of their search was a ribozyme that could carry out an aldol reaction about 4,300-fold faster than the reaction would occur in solution. This catalyst was difficult to find: there was only one sequence that could carry out the

reaction efficiently among the 10¹⁵ starting RNAs. This unique large RNA had a major nucleotide tract that could be deleted without affecting the aldol reaction (Fig. 1). This is to be expected given the single-pass nature of the experiment — had an aldol ribozyme evolved naturally, this 'junk' would probably have been whittled away by the selection.

The new ribozyme absolutely requires zinc ions, which are also uniquely sufficient for the RNA's activity (magnesium, barium, copper, manganese and many other metals will not, on their own, activate ribozymes). This is notable because, of the two common types of protein enzyme that carry out aldol reactions in modern cells, one uses zinc as a catalytic element. So the RNA found by Fusz *et al.* may resemble an existing protein in the way it polarizes its reactants to speed up the aldol reaction.

Finding an aldol ribozyme is provocative because an RNA world would have required some form of energy metabolism; energy must be captured to maintain complex and unstable molecules such as RNA. Rudimentary catalysts for the replication⁸ of RNA or the synthesis of proteins⁹ were isolated some time ago. But very few RNAs, either natural or recently winnowed out by artificial selection, supply clues to how energy might have been harvested in an RNA world. So there is a void even in purely imaginative accounts of the RNA world, which these findings help to fill.

The reverse aldol (retroaldol) reaction (Fig. 1) is interesting in this light because although not itself an energy-harvesting reaction, it enhances the energy yield from hexose sugars such as glucose during glycolysis. In today's biology, it is closely coupled to reactions that yield ATP, the energy currency for most cellular work. The new reaction is thus a potentially useful part, if not the heart, of a possible metabolic RNA engine that would eat carbohydrates.

Michael Yarus is in the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA. e-mail: yarus@colorado.edu

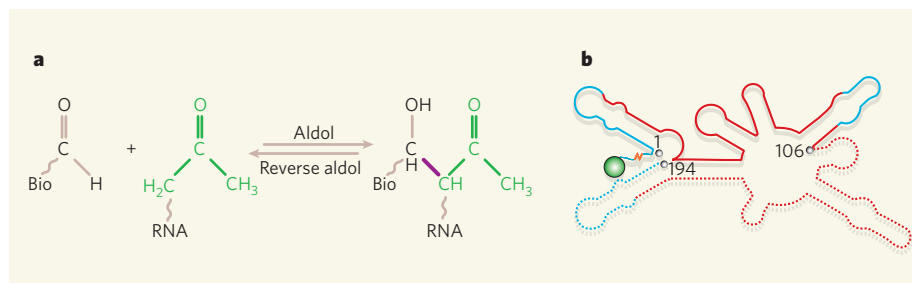


Figure 1 | Resurrecting an aldol ribozyme. **a**, Fusz *et al.*¹ tested a vast array of random-sequence RNAs to see whether any of them could speed up this predicted variant of the aldol reaction. The new carbon–carbon bond produced is shown in purple, and the reactive 'ketone' substrate group is shown in green. Bio is biotin, a chemical tag that allows the molecule to be easily separated from the reaction mixture. **b**, The selected ribozyme. The numbers represent nucleotides counting from the 5' end. Blue shows fixed sequences and red shows the sequences that were randomized. Broken lines show those nucleotides that can be deleted without destroying the molecule's activity. The green sphere is the ketone substrate, and the orange zigzag shows the light-sensitive link, used to ensure that biotin addition has occurred at the ketone, and not elsewhere.

1. Fusz, S., Eisenführ, A., Srivatsan, S. G., Heckel, A. & Famulok, M. *Chem. Biol.* **12**, 941–950 (2005).
2. Helm, M., Petermeier, M., Ge, B., Fiammengio, R. & Jäschke, A. *J. Am. Chem. Soc.* **127**, 10492–10493 (2005).
3. Wilson, D. S. & Szostak, J. W. *Annu. Rev. Biochem.* **68**, 611–647 (1999).
4. Lilley, D. M. J. *Curr. Opin. Struct. Biol.* **15**, 313–323 (2005).
5. Yarus, M. & Knight, R. in *The Genetic Code and the Origin of Life* (ed. de Pouplana, L.) 75–91 (Landes Bioscience, Georgetown, Texas, 2004).
6. Yarus, M. *Annu. Rev. Biochem.* **74**, 179–198 (2005).
7. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. *Nature* **389**, 54–57 (1997).
8. Lawrence, M. S. & Bartel, D. P. *RNA* **11**, 1173–1180 (2005).
9. Yarus, M. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 207–215 (2001).

Correction

In "Ecology: Roots of stability" by Peter D. Moore (*Nature* **437**, 959–961; 2005), the reference to the main paper discussed was incorrect. The correct reference is Kahmen, A., Perner, J. & Buchmann, N. *Funct. Ecol.* **19**, 594–601 (2005).

OBITUARY

Richard Doll (1912–2005)

Epidemiologist extraordinary.

Only 60 years ago, smoking was generally regarded as innocuous. Now it is recognized as the main cause of lung cancer, the world's commonest cancer. More than any other person, Richard Doll was responsible for this remarkable transformation in attitude, as well as for the now widely accepted view that much cancer (not just of the lung) is, at least in principle, avoidable. Doll, who died on 24 July 2005 aged 92, was the world's foremost epidemiologist and had a wide influence on the progress of the subject.

Doll was born on 28 October 1912, the son of a London general practitioner. He might have pursued his strong interest in mathematics, but at the last moment decided to follow his father into medicine, qualifying at St Thomas' Hospital in 1937. His epidemiological career began after war service, with work on the causes of peptic ulcers that attracted the attention of Tony Bradford Hill. At this time many considered that the marked rise in lung cancer deaths between the two world wars was simply due to improved diagnosis. In 1948, however, Percy Stocks and Ernest Kennaway succeeded in persuading the Medical Research Council (MRC) to investigate the issue.

For this, the MRC turned to Hill, who recruited Doll, and together they obtained detailed smoking histories from a large number of patients with and without the disease. In 1950 they reported the findings of this 'case-control' study. The rarity of non-smokers and the relative excess of heavy smokers among the lung-cancer cases, compared with the controls, convinced them that cigarette smoking was "a cause, and an important cause" of the disease.

This was not, in fact, the first such study, for earlier in the same year similar findings were reported in the United States by Ernst Wynder and Ewart Graham. Doll and Hill, however, went on to collect details of smoking habits from 40,000 British doctors, and to ascertain the causes of death of those who died. Their earlier findings were confirmed, including the dose-response relationship (the more cigarettes smoked, the greater the lung cancer risk). But this 'cohort' approach was also able to reveal the wide range of other smoking-related diseases; in the recent 50-year follow-up, published with Richard Peto, he showed that one-half of persistent smokers were estimated to die as a result of the habit. The long interval from starting smoking before its major effects on mortality appear was another notable finding.

This study has continued and is unique in

its regular updating of the smoking habits of participants that has allowed recognition of the (unexpected) benefits of stopping smoking. The strong dose-response relation between lung cancer and cigarette smoking, the high standard of the design and conduct of the study, and the balanced assessment of its findings in many papers — all of these played a part in making a convincing case about the perils of smoking. In consequence, habits have changed. For example, in Britain the proportion of men who smoke has fallen from 80% a few decades ago to less than 30%.

Doll's work went far beyond smoking, however. He was involved in early investigations of the carcinogenic effects of ionizing radiation; in 1957 with Michael Court Brown he followed the fates of 14,000 patients with ankylosing spondylitis who were treated with radiation. This cohort study brought independent confirmation of studies of Japanese A-bomb survivors that radiation could cause leukaemia, and it has been a major source of data on the dose-response relation of radiation and cancer.

In another project, British radiologists were investigated because of their repeated exposure to low doses of radiation, the only group with excess cancer being those with long practice in the early twentieth century when high cumulative doses were likely as a result of the relative lack of protective procedures. Later Doll studied many aspects of exposure to low levels of radiation, both ionizing (for example radon and its relationship to lung cancer) and non-ionizing (electromagnetic fields from power lines and childhood leukaemia).

Doll was an author of the first compendium of worldwide data on cancer incidence, and was the first to recognize that each cancer that was common in one part of the world was rare in another — and for reasons that are not primarily genetic, but largely extrinsic. In 1954 with Peter Armitage, long before the relevant advances in molecular biology, he adduced evidence for the multi-stage nature of carcinogenesis. In 1955 he was also the first to show a significant excess of lung cancer among asbestos workers and later, with Martin Vessey and others, he studied the side-effects of oral contraceptives. He continued a clinical attachment at the Central Middlesex Hospital until 1969, conducting therapeutic trials on peptic ulcers, including the first demonstration of the efficacy of both liquorice extract and carbenoxolone.

In 1969 Doll left the MRC Statistical



Research Unit in London (of which he had been director since 1961) to become Regius Professor of Medicine at Oxford, an unexpected appointment that brought added attention to the subject of epidemiology. The range of his epidemiological activities was extraordinarily varied. In his career he published over 500 papers, reputedly without the irritation of writing a single grant application (he was, incidentally, saddened that new government rules will prevent certain harmless but valuable research into public health, like some of his own). Late in life he participated in randomized trials, and meta-analyses of trials, of treatments for breast cancer and vascular disease, as well as in collaborative meta-analyses of risk factors for breast cancer.

Doll's approach to his work was exemplary. He examined data with extraordinary detachment, even when surrounded by intense preconceptions. The more closely people worked with him, the more respect they had for him. Highly (but unobtrusively) organized, his concentration, efficiency and memory were formidable — as was his capacity for work, which was unaffected by long-haul flights or age.

He is commemorated in Oxford by Green College, which he established, and by the two research units he helped to found. Both of these are now housed in the new Richard Doll Building.

Leo Kinlen

Leo Kinlen is in the Cancer Epidemiology Unit, Richard Doll Building, Roosevelt Drive, Headington, Oxford OX3 7LF, UK. e-mail: leo.kinlen@dphpc.ox.ac.uk

R. JUDGES/REX FEATURES

BRIEF COMMUNICATIONS

Crowd synchrony on the Millennium Bridge

Footbridges start to sway when packed with pedestrians falling into step with their vibrations.

Soon after the crowd streamed on to London's Millennium Bridge on the day it opened, the bridge started to sway from side to side: many pedestrians fell spontaneously into step with the bridge's vibrations, inadvertently amplifying them. Here we model this unexpected and now notorious phenomenon — which was not due to the bridge's innovative design as was first thought — by adapting ideas originally developed to describe the collective synchronization of biological oscillators such as neurons and fireflies. Our approach should help engineers to estimate the damping needed to stabilize other exceptionally crowded footbridges against synchronous lateral excitation by pedestrians.

Existing theories^{1–6} of what happened on the bridge's opening day focus on the wobbling of the bridge but have not addressed the crowd-synchronization dynamics. In our approach, wobbling and synchrony are inseparable. They emerge together, as dual aspects of a single instability mechanism, once the crowd reaches a critical size.

We model the bridge as a weakly damped and driven harmonic oscillator

$$M \frac{d^2 X}{dt^2} + B \frac{dX}{dt} + KX = G \sum_{i=1}^N \sin \Theta_i \quad (1)$$

where $X(t)$ is the displacement of the relevant lateral mode, and M , B and K are its modal mass, damping and stiffness, respectively (Fig. 1). Each pedestrian $i = 1, \dots, N$ imparts an alternating sideways force $G \sin \Theta_i$ to the bridge, where G is the maximum force and the phase $\Theta_i(t)$ increases by 2π during a full left/right walking cycle.

The bridge's movement, in turn, is assumed to alter each pedestrian's gait according to

$$\frac{d\Theta_i}{dt} = \Omega_i + C A \sin(\Psi - \Theta_i + \alpha) \quad (2)$$

where the frequencies Ω_i are randomly distributed with a density $P(\Omega)$, reflecting the diversity of natural footfall rates across the population; C quantifies pedestrians' sensitivity to bridge vibrations of amplitude $A(t)$ and phase $\Psi(t)$ (defined such that $X = A \sin \Psi$, $dX/dt = \Omega_0 A \cos \Psi$, where $\Omega_0 = \sqrt{K/M}$ is the

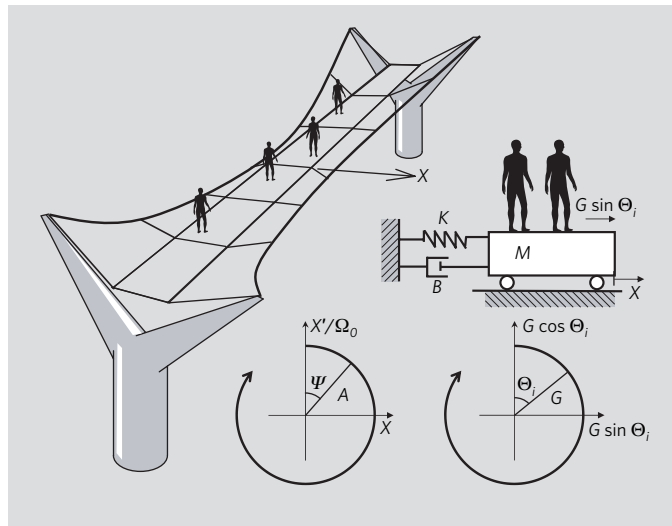


Figure 1 | Effect of pedestrian crowding on London's Millennium Bridge. The resonant lateral mode of vibration of the bridge (left) can be represented by a mass-spring-damper system (top, right). The angular phases (bottom) for the bridge displacement X (left) and the individual pedestrian forces, $G \sin \Theta_i$ (right), are indicated (see text for definitions of variables).

bridge's resonant frequency); and α is a phase lag parameter. Equation (2) is hypothetical but testable (see supplementary information), and is consistent with the known response of other biological rhythms to weak periodic stimuli^{7,8}.

To illustrate the dynamics inherent in the model, Fig. 2 shows a simulation of a controlled experiment² performed on the Millennium Bridge. As more and more people walk on to the deck (Fig. 2a), there is no hint of

Figure 2 | Simulated outbreak of wobbling and crowd synchronization on the Millennium Bridge.

a, Number of pedestrians, N , slowly increasing stepwise, as used in a diagnostic wobble test² on the bridge's north span in December 2000. **b**, Predicted amplitude $A = (X^2 + (\Omega_0^{-1} dX/dt)^2)^{1/2}$ of the bridge's resulting lateral vibrations. **c**, Predicted degree of phase coherence among the pedestrians, as measured by the order parameter^{9,10} $R = N^{-1} |\sum_{i=1}^N \exp(i\Theta_i)|$. For small crowds, walkers are desynchronized and randomly phased: hence R fluctuates and decays as $R \approx N^{-1/2}$. At a critical crowd size (dashed lines in **a–c**), the bridge starts to sway and the crowd starts to synchronize, with each process pumping the other in a positive feedback loop; R and A follow parallel time-courses (**b, c**), maintaining an almost constant ratio. This predicts the empirical observation⁴ that the crowd drives the bridge with a force proportional to the bridge's velocity (see supplementary information for details).

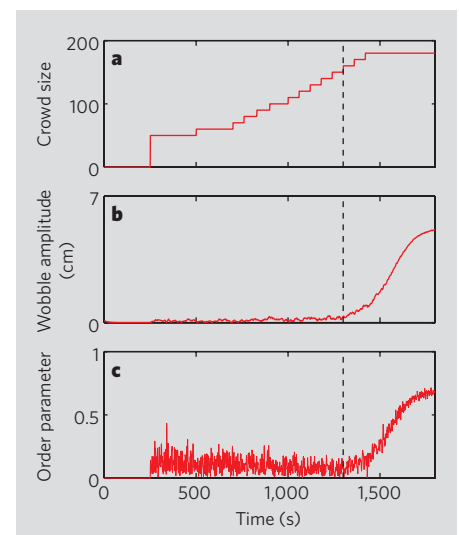
instability until the crowd reaches a critical size, N_c , after which wobbling and synchrony erupt simultaneously (Fig. 2b, c). We can calculate N_c analytically, using methods^{8–10} created to study large systems of biological oscillators (see supplementary information). To take the simplest case, suppose $\alpha = \pi/2$ and $P(\Omega)$ is symmetrical about Ω_0 (also a 'worst case' for the bridge, in the sense that pedestrians then drive it most efficiently). We find

$$N_c = \frac{4s}{\pi} \left(\frac{K}{GC P(\Omega_0)} \right) \quad (3)$$

where $s = B/\sqrt{4MK}$ is the damping ratio. All the parameters have known values, except for C . Comparing our simulations with data obtained from crowd tests on the Millennium Bridge², we estimate $C \approx 16 \text{ m}^{-1} \text{ s}^{-1}$. Then, with no fur-

ther adjustable parameters, the model predicts the synchronization timescale and the characteristic amplitude of the wobbles shown in Fig. 2 and in the actual experiments^{1,2}. It also accounts for the previously unexplained empirical observation⁴ that the excitation force generated by the crowd grows linearly with the bridge's amplitude (for calculations, see supplementary information).

By generalizing ideas that were developed in



mathematical biology, we have provided a unified picture of what happened on the Millennium Bridge five years ago, both for the bridge vibrations and the crowd dynamics. The approach suggested here may also prove useful for estimating the damping needed to safeguard other bridges, present and future, against synchronous lateral excitation by pedestrians.

Steven H. Strogatz*, **Daniel M. Abrams***,
Allan McRobie†, **Bruno Eckhardt‡**, **Edward Ott‡**

*Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853-1503, USA

e-mail: strogatz@cornell.edu

†Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

‡University of Maryland, College Park, Maryland 20742, USA

SFachbereich Physik, Philipps-Universität Marburg, 35032 Marburg, Germany

1. Dallard, P. *et al. Struct. Eng.* **79**, 17–33 (2001).
2. Dallard, P. *et al. J. Bridge Eng.* **6**, 412–417 (2001).
3. McRobie, A., Morgenthal, G., Lasenby, J. & Ringer, M. *Proc. Inst. Civ. Eng. Bridge Eng.* **156**, 71–79 (2003).
4. Roberts, T. M. *Proc. Inst. Civ. Eng. Bridge Eng.* **156**, 155–160 (2003).
5. Newland, D. E. *Proc. Inst. Mech. Eng.* **218**, 477–492 (2004).
6. Nakamura, S. *J. Struct. Eng.* **130**, 32–37 (2004).
7. Glass, L. & Mackey, M. C. *From Clocks to Chaos: The Rhythms of Life* (Princeton Univ. Press, Princeton, 1988).
8. Winfree, A. T. *The Geometry of Biological Time* 2nd edn (Springer, New York, 2001).
9. Kuramoto, Y. *Chemical Oscillations, Waves and Turbulence* (Springer, Berlin, 1984).
10. Strogatz, S. H. *Physica D* **143**, 1–20 (2000).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/43843a

NANOSCALE HYDRODYNAMICS

Enhanced flow in carbon nanotubes

Nanoscale structures that could mimic the selective transport and extraordinarily fast flow possible in biological cellular channels would have a wide range of potential applications. Here we show that liquid flow through a membrane composed of an array of aligned carbon nanotubes is four to five orders of magnitude faster than would be predicted from conventional fluid-flow theory. This high fluid velocity results from an almost frictionless interface at the carbon-nanotube wall.

Biological channels act as chemically selective gatekeepers and have protein walls that allow extremely rapid transit¹. Nanometre-scale pores with chemical selectivity have been prepared^{2,3} but fluid flow through them is slow: this limitation is predicted by the Hagen–Poiseuille equation and is because conventional laminar flow has zero fluid velocity at the pore walls.

In theory, the flow of molecules inside carbon nanotubes could be much faster. Water should be able to flow fast through hydrophobic single-walled carbon nanotubes because the process creates ordered hydrogen bonds between the water molecules⁴. Ordered hydrogen bonds between water molecules and the weak attraction between the water and smooth carbon-

nanotube graphite sheets should then result in almost frictionless and very rapid flow⁴. If a theoretical volume rate comparable to that of the protein channel aquaporin-1 (ref. 4) is divided by the carbon-nanotube cross-sectional area, the expected water flow velocity is about 90 cm s⁻¹. Fast flow velocities are also predicted just from the frictionless nature of the carbon-nanotube walls⁵ and from the rapid diffusion of hydrocarbons^{6,7}.

To realize these high flow velocities, we used a freshly fabricated membrane consisting of aligned multiwalled carbon nanotubes, with graphitic inner cores (diameter about 7 nm) and a high area density (5 × 10¹⁰ per cm²), crossing a solid polystyrene film⁸. We measured the flow of water and a variety of solvents through this membrane at about 1 atm applied pressure (Table 1). In a control experiment, we verified that no macroscopic defects were present in the membrane and determined the available pore area (see supplementary information).

We found that the flow rates are four to five orders of magnitude faster than conventional fluid flow would predict through pores of 7 nm diameter. Contrary to predictions based on hydrodynamics, the flow rate does not decrease with increased viscosity (compare hexane and water in Table 1). The results also indicate that flow velocity, when adjusted for differences in viscosity, increases for more hydrophilic fluids. The flow of hydrogen-bonded fluids decreases after a few minutes, but this does not occur with alkanes or aqueous solutions of potassium chloride. Reduction in the flow of

associated liquids (water and alcohols) with time can be attributed to flow-induced solvent ordering or the formation of bubbles (our unpublished results).

We conclude that these high fluid velocities are possible because of a frictionless surface at the carbon-nanotube wall. This result could be explained in conventional terms of slip lengths, which are remarkably long. The slip length is an extrapolation of the extra pore radius required to give zero velocity at a hypothetical pore wall (the boundary condition for conventional materials). The observed slip lengths (3–70 μm) are much longer than the pore radius (3.5 nm) that is consistent with a nearly frictionless interface. The slip length decreases as solvents become more hydrophobic (Table 1), which indicates stronger interaction with the carbon-nanotube wall. The observed flow velocities for water (10–44 cm s⁻¹) are close to the extrapolated rate predicted for water flow through single-walled carbon nanotubes (about 90 cm s⁻¹)⁴. Butane flows through carbon nanotubes at about 26 cm s⁻¹ (ref. 6), which is consistent with our measurement for hexane.

These results show that the speed of fluid flow through the aligned carbon-nanotube membrane approaches that through biological channels. The membrane fabrication is scalable to large areas, which could be useful industrially for chemical separations; chemical functionality is near the core entrance⁹ and each side of the membrane can be independently modified with different functional groups¹⁰. These advantages also make the aligned carbon-nanotube membrane a promising mimic of protein channels for transdermal drug delivery and selective chemical sensing.

Mainak Majumder*, **Nitin Chopra***,
Rodney Andrews†, **Bruce J. Hinds***

*Chemical and Materials Engineering Department, University of Kentucky, Lexington, Kentucky 40506, USA

e-mail: bjhinds@engr.uky.edu

†Center for Applied Energy Research, Lexington, Kentucky 40511, USA

1. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1984).
2. Jirage, K. B., Hulteen, J. C. & Martin, C. R. *Science* **278**, 655–658 (1997).
3. Klein, E. J. *Membr. Sci.* **179**, 1–27 (2000).
4. Hummer, G., Rasaiah, J. C. & Noworyta, J. P. *Nature* **414**, 188–190 (2001).
5. Sokhan, V. P., Nicholson, D. & Quirke, N. *J. Chem. Phys.* **117**, 8531–8539 (2002).
6. Skoulidas, A. I., Ackerman, D. M., Johnson, J. K. & Sholl, D. S. *Phys. Rev. Lett.* **89**, 185901 (2002).
7. Mao, Z. & Sinnott, S. B. *J. Phys. Chem. B* **105**, 6916–6924 (2001).
8. Hinds, B. J. *et al. Science* **303**, 62–65 (2004).
9. Majumder, M., Chopra, N. & Hinds, B. J. *J. Am. Chem. Soc.* **127**, 9062–9067 (2005).
10. Chopra, N., Majumder, M. & Hinds, B. J. *Adv. Funct. Mater.* **15**, 858–864 (2005).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared online.
doi:10.1038/43844a

Table 1 | Pressure-driven flow through aligned MWCNT membrane

Liquid	Initial permeability*	Observed flow velocity†	Expected flow velocity‡	Slip length (nm)
Water	0.58	25	0.00057	54
	1.01	43.9	0.00057	68
	0.72	9.5	0.00015	39
Ethanol	0.35	4.5	0.00014	28
iso-Propanol	0.088	1.12	0.00077	13
Hexane	0.44	5.6	0.00052	9.5
Decane	0.053	0.67	0.00017	3.4

MWCNT, multiwalled carbon nanotube. For details of methods, see supplementary information. *Units, cm³ per cm² min bar. †Flow velocities in cm s⁻¹ at 1 bar. ‡Expected flow velocity is that predicted from conventional flow.

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

ECOLOGY

Is speciation driven by species diversity?

Arising from: B. C. Emerson & N. Kolm *Nature* 434, 1015–1017 (2005)

Emerson and Kolm¹ show that the proportion of species endemic to an island is positively related to its species richness and, assuming that endemism indexes speciation rate, they infer that greater species diversity accelerates diversification. Here we demonstrate that the same correlation between species richness and percentage endemism can arise even if within-island speciation is negligible, particularly when both endemism and species richness depend on attributes of islands (such as area) that influence the average age of resident populations. Island biogeography theory indicates that, where the average time to extinction is relatively long, diversity increases through

colonization, irrespective of whether new species are formed²; at the same time, islands on which populations persist for longer accumulate more endemic species as local populations differentiate and populations on neighbouring islands become extinct^{3,4}. We therefore suggest that species richness and endemism are correlated fortuitously owing to their mutual dependence on the life spans of populations on islands, which is unrelated to speciation itself.

If the scenario we propose is correct, islands richer in species and with higher endemism would also have older populations. This prediction is supported by data for birds in the

West Indies, a system in which the occurrence of within-island speciation is an extremely rare event: the proportion of endemic species increases linearly with species richness, and species-poor islands with few endemics have populations that are younger on average than those of species-rich islands with higher endemism (Fig. 1). We argue, therefore, that the correlation between species richness and endemism does not imply a causal relationship between these two variables, but rather that they respond in parallel to the effect that area and other island attributes have on population persistence times. Accordingly, the slope of the species–area relationship is much steeper for old endemics that are largely absent from small, species-poor islands, than for recent colonists, which are similarly diverse on all islands⁵. The fact that large (that is, species-rich) islands have populations with longer persistence times, as indicated by genetic-distance data, is consistent with island biogeography theory for systems where species richness is influenced primarily by rate of extinction⁶.

Our reasoning implies that per cent endemism will often be an inappropriate surrogate for speciation rate because it is strongly influenced by differentiation between island populations and extinction (see also ref. 7). Our analysis also contradicts Emerson and Kolm's assertion that extinction on neighbouring islands would inflate the proportion of endemic species more strongly for species-poor islands. This assertion follows from their argument that high species diversity drives a high rate of extinction. Although total extinctions per unit time rise as diversity increases on an island over time², the rate of extinction per species, hence the average species life span, is relatively insensitive to species richness. Moreover, theory² and observation (Fig. 1) indicate that the rate of extinction per species is inversely related to species richness in comparisons between islands.

The data shown in Fig. 1 support our proposed scenario but disagree with a prediction of Emerson and Kolm's hypothesis — namely that, if speciation rate accelerates as local diversity increases, the rate of turnover under speciation–extinction equilibrium should be higher and the average duration of populations should therefore be shorter on more diverse islands. Some insular endemics are of recent origin^{8,9}, but whether they arose on islands on which populations persist for relatively longer (because of larger island size, increased habitat diversity, milder weather or other factors) can only be addressed with data on persistence times for multiple lineages and

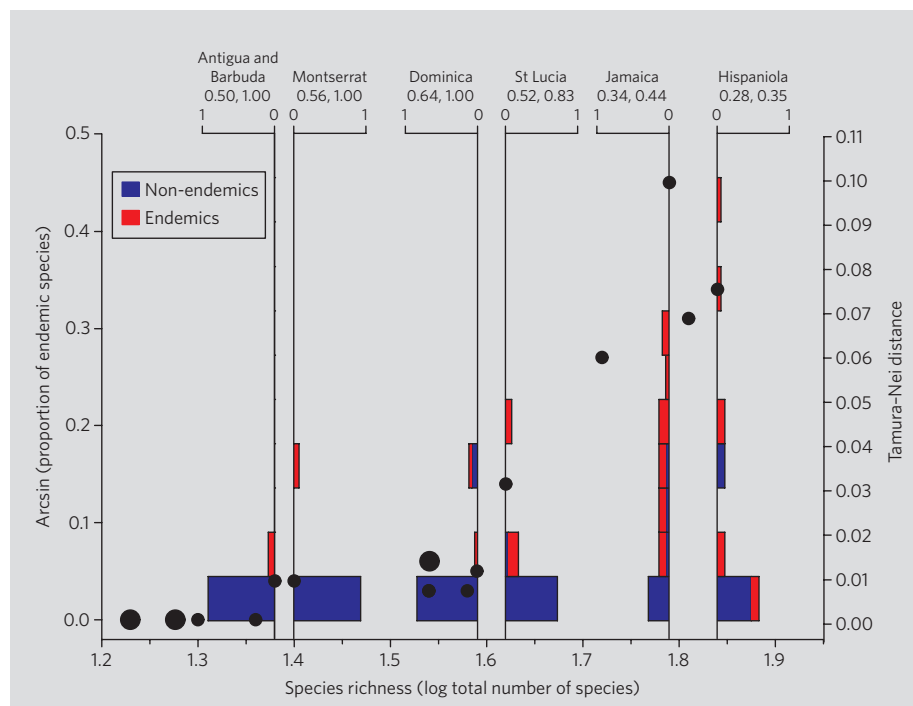


Figure 1 | Relationship between species richness, endemism and persistence times of populations of resident land birds on Caribbean islands. Black dots indicate values of species richness and endemism; larger dots indicate values shared by two different islands. Histograms show relative frequency distribution of persistence times of non-endemic and endemic populations on six islands (Antigua and Barbuda are considered as a single island), measured as mitochondrial DNA Tamura–Nei distance to the node uniting them with their nearest conspecific population or sister taxon. Numbers above the histograms are the proportion of the total species and of endemic species for which we obtained genetic distance data for each island. The vertical axes of histograms cross the dots that correspond to the species richness and endemism of the islands for which they portray genetic distance data. As with plants in the Hawaiian archipelago¹, species richness and island area are highly correlated ($r = 0.928$) and their effects on endemism cannot be separated. When area is excluded from the analysis, as in one of the examples in ref. 1, species richness is the only variable that explains endemism in a forward multiple regression that includes maximum elevation and distance to the nearest island as additional independent variables ($r^2 = 0.723$, $F = 41.671$, $P < 0.0001$, $\beta = 0.593 \pm 0.09$ s.e., $t = 6.46$). Populations on species-rich islands are significantly older than those on species-poor islands (Kruskal–Wallis $\chi^2 = 38.1$, d.f. = 5, $P < 0.0001$), and endemics are older than non-endemics on each of the six islands (Kruskal–Wallis χ^2 , 4.9–13.6; d.f. = 1; P , 0.0002–0.027). Data on mitochondrial genetic differentiation are for 113 populations of 60 bird species on six islands with varying numbers of species and degrees of endemism. Further details are available from the authors (www.umsl.edu/~cdc35b).

islands. In the absence of this information, and of meaningful estimates of intervals between speciation events, Emerson and Kolm's approach to the idea that species diversity might drive diversification is inconclusive. This hypothesis might be plausible in systems where speciation events take place readily within islands, including those described by Emerson and Kolm. However, the influence of diversity on species formation can be properly addressed only by considering variation in per-lineage speciation rate, estimated from phylogenetic reconstructions^{10,11}, across areas with varying species richness.

Carlos Daniel Cadena*, **Robert E. Ricklefs***,
Iván Jiménez†, **Eldredge Bermingham‡**

*Department of Biology, University of Missouri,
St Louis, Missouri 63121, USA
e-mail: cdc35b@umsl.edu

†Center for Conservation and Sustainable
Development, Missouri Botanical Garden,

PO Box 299, St Louis, Missouri 63166, USA
‡Smithsonian Tropical Research Institute,
Box 2072, Balboa, Republic of Panama

1. Emerson, B. C. & Kolm, N. *Nature* **434**, 1015–1017 (2005).
2. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 1967).
3. Mayr, E. *Science* **150**, 1587–1588 (1965).
4. Losos, J. B. & Schluter, D. *Nature* **408**, 847–850 (2000).
5. Ricklefs, R. E. & Bermingham, E. *Am. Nat.* **163**, 227–239 (2004).
6. Johnson, K. P., Adler, F. R. & Cherry, J. L. *Evolution* **54**, 387–396 (2000).
7. Goldberg, E. E., Roy, K., Lande, R. & Jablonski, D. *Am. Nat.* **165**, 623–633 (2005).
8. Baldwin, B. G. & Sanderson, M. J. *Proc. Natl Acad. Sci. USA* **95**, 9402–9406 (1998).
9. Emerson, B. C. & Oromí, P. *Evolution* **59**, 586–598 (2005).
10. Nee, S., Mooers, A. Ø. & Harvey, P. H. *Proc. Natl Acad. Sci. USA* **89**, 8322–8326 (1992).
11. Pybus, O. G. & Harvey, P. H. *Proc. R. Soc. Lond. B* **267**, 2267–2272 (2000).

doi:10.1038/nature04308

ECOLOGY

Emerson & Kolm reply

Replying to: C. D. Cadena, R. E. Ricklefs, I. Jiménez & E. Bermingham *Nature* **438**,
doi:10.1038/nature04308 (2005)

Cadena *et al.*¹ question our conclusion that species diversity can positively influence speciation rate on the basis of their analysis of a data set for West Indian land birds, in which an additional variable is added — lineage age. Here we clarify our hypothesis and show why their system is not suitable for testing whether species diversity can drive speciation.

Cadena *et al.* find that lineage age is correlated with species diversity and per cent endemism. However, as they point out, a fourth variable, island area, is also strongly collinear with species diversity and endemism, and so with lineage age. A similar collinearity between island area and species diversity occurs for one of our four analyses, Hawaiian plants, and we recognized the difficulty of disentangling the effect of these two variables.

For West Indian land birds, island area is likely to be the causative agent for the observed levels of endemism, perhaps because bigger islands contain older species assemblages that have had more time to accumulate endemics, as Cadena *et al.* suggest. Hence, if island area and not species diversity is driving diversification of the avifauna of the West Indies — a likely scenario, given its non-equilibrium state with an imbalance between colonization and extinction² — Cadena *et al.* risk comparing apples with oranges.

MacArthur and Wilson's classic theory³ is traditionally interpreted in terms of colonization and extinction, but we pointed out that it also makes predictions for speciation⁴. Even in the simplest scenario of anagenetic speciation only, the theory of island biogeog-

raphy predicts that, all other things being equal, islands with more species will have a greater proportion of endemics (Fig. 1a). Here the proportion of endemics does provide an index of speciation (and extinction) rate, contrary to the assertion of Cadena *et al.*¹. Similarly, all other things being equal, larger islands are expected to have a smaller proportion of endemics than smaller islands (Fig. 1b). This discrepancy with the results of Cadena *et al.* is consistent with the non-equilibrium nature of the West Indian avifauna, where colonization does not yet seem to be balanced by extinction². Our predictions are expected for the systems we used⁴, where there is a balance between colonization, speciation and extinction (as seen in Canary Island arthropods⁵, where island area and age are not positively related to the proportion of endemics).

Cadena *et al.* also point out that predictions regarding lineage age made from the theory of island biogeography can be tested by molecular phylogenetics. However, lineage age is better determined using age estimates for the most recent common ancestor of monophyletic groups within islands (see ref. 6, for example). This provides a conservative minimum age estimate to allow for extinctions between the most recent common ancestor of island species and that connecting an island clade to a sister lineage on another land mass⁷. Extinctions between the nodes that involve taxa outside the island will inflate the true population age. Hence, the lineage age estimation of Cadena *et al.* may be misleading, particularly in view of taxon-cycling theory as applied to the West Indian avifauna, which supports a relationship between endemism and sister lineage extinction⁸.

We agree that our theory of how species richness drives diversification may be less important in systems that are not under equilibrium conditions. But the analysis by Cadena *et al.* does little to bring into question our conclusion that species diversity may be an important driver of speciation.

Brent C. Emerson*, **Niclas Kolm†**

*School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

e-mail: b.emerson@uea.ac.uk

†Present address: Institute of Evolutionary Biology, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, UK

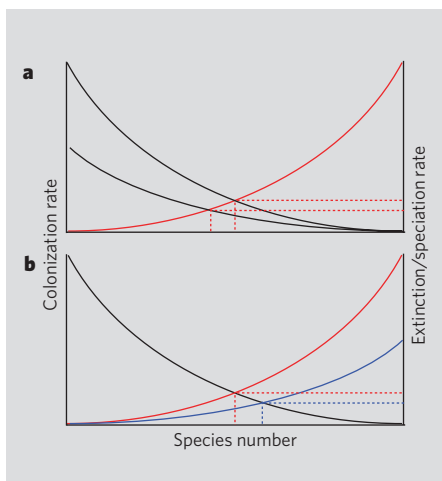


Figure 1 | The theory of island biogeography and speciation rate. **a**, The number of species on an island is a balance between the arrival of colonizing species to an island (black lines) and extinction (red line)³. For two islands of the same size but with differing colonization rates, the island receiving more colonists is expected to contain more species. As the same factors that influence extinction rate also influence speciation rate⁴, the model can predict speciation rate: the island with more species should have a greater proportion of endemics because of a comparatively higher speciation rate. **b**, For two islands with a similar colonization rate but having different numbers of species because of their different sizes, the smaller island (red line) should have a greater proportion of endemic species than the larger island (blue line).

1. Cadena, C. D., Ricklefs, R. E., Jiménez, I. & Bermingham, E. *Nature* **438**, doi:10.1038/nature04308 (2005).
2. Ricklefs, R. E. & Bermingham, E. *Science* **294**, 1522–1524 (2001).
3. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 1967).
4. Emerson, B. C. & Kolm, N. *Nature* **434**, 1015–1017 (2005).
5. Emerson, B. C. & Oromí, P. *Evolution* **59**, 586–598 (2005).
6. Emerson, B. C. *Anim. Biodiv. Cons.* **26**, 9–20 (2003).
7. Emerson, B. C. *Mol. Ecol.* **11**, 951–966 (2002).
8. Ricklefs, R. E. & Bermingham, E. *Ostrich* **70**, 49–59 (1999).

doi:10.1038/nature04309

Tracing the first stars with fluctuations of the cosmic infrared background

A. Kashlinsky^{1,2}, R. G. Arendt^{1,2}, J. Mather^{1,3} & S. H. Moseley^{1,3}

The deepest space- and ground-based observations find metal-enriched galaxies at cosmic times when the Universe was less than 1 Gyr old. These stellar populations had to be preceded by the metal-free first stars, known as 'population III'. Recent cosmic microwave background polarization measurements indicate that stars started forming early—when the Universe was ≤ 200 Myr old. It is now thought that population III stars were significantly more massive than the present metal-rich stellar populations. Although such sources will not be individually detectable by existing or planned telescopes, they would have produced significant cosmic infrared background radiation in the near-infrared, whose fluctuations reflect the conditions in the primordial density field. Here we report a measurement of diffuse flux fluctuations after removing foreground stars and galaxies. The anisotropies exceed the instrument noise and the more local foregrounds; they can be attributed to emission from population III stars, at an era dominated by these objects.

The cosmic infrared background (CIB) is generated by emission from luminous objects during the entire history of the Universe, including epochs at which discrete objects are inaccessible to current telescopic studies^{1,2}. With new powerful telescopes, individual galaxies are now found out to redshifts of $z \gtrsim 5$ –7, but the period preceding that of the galaxies seen in the Hubble ultra-deep field remains largely unexplored. These ultra-deep field galaxies and even the highest- z quasars ($z \gtrsim 6.5$) appear to consist of 'ordinary' metal-enriched population I and II stars, suggesting that the metal-free stars of population III existed at still earlier epochs. Large-scale polarization of the cosmic microwave background (CMB)³ suggests that first stars formed early, at $z \approx 20$.

Current theory predicts that population III objects were very massive stars, with mass greater than 100 solar masses ($100M_{\odot}$; refs 4–6). They should have produced a significant diffuse background⁷ with an intensity almost independent of the details of their mass function⁸. Because much of the emission at wavelengths longer than the rest-frame Lyman limit from these epochs is now shifted into the near-infrared (NIR), these stars could be responsible for producing much, or all, of the observed NIR CIB excess over that from normal galaxy populations (see detailed review in ref. 2, and refs 8–13). Two groups^{8,13} have recently suggested that this emission should have a distinct angular spectrum of anisotropies, which could be measured if the contributions from the ordinary (metal-rich) galaxies and foregrounds could be isolated and removed, and provide an indication of the era made predominantly of the massive population III stars.

Measuring CIB anisotropies from objects at high z is difficult, because the spatial fluctuations are small and can be hidden by the contributions of ordinary low- z galaxies as well as instrument noise and systematic errors in the data. Previous attempts to measure the structure of the CIB in the NIR on the degree scale with COBE/DIRBE¹⁴ and IRTS/NIRS¹⁵ were limited in sensitivity because of the remaining contributions from brighter galaxies in the large beams. Analysis of 2MASS data at 1.25, 1.65 and 2.2 μm with ~ 2 arcsec resolution^{16,17} allowed removal of foreground galaxies to a K magnitude of 19–19.5 (AB magnitude $m_{\text{AB}} \approx 21$ or 15 μJy), and led to measurements of CIB fluctuations at 1.25 to 2.2 μm at sub-arcminute

angular scales. These studies reported fluctuations in excess of that expected from the observed galaxy populations, although their accurate interpretation in terms of high- z contributions is difficult owing to foreground galaxies and non-optimal angular scales.

We used data from deep exposure data obtained with Spitzer/IRAC^{18,19} in four channels (channels 1–4 correspond to wavelengths of 3.6, 4.5, 5.8 and 8 μm , respectively) in attempting to uncover this signal. We find significant CIB anisotropies after subtracting galaxies substantially fainter than was possible in prior studies, that is, down to $m_{\text{AB}} \approx 22$ –25. The angular power spectrum of the anisotropies is significantly different from that expected from Solar System and Galactic sources, its amplitude is much larger than what is expected from the remaining faint galaxies, and can reasonably be attributed to the diffuse light from the population III era.

Assembly and reduction of data sets

The primary data set that we used here is the deep IRAC observation of a $\sim 12' \times 6'$ region around the quasi-stellar object (QSO) HS 1700+6416 obtained by the Spitzer instrument team^{18,20}. In addition, the data from two auxiliary fields with shallower exposures (HZF and EGS) were analysed. Relevant data characterizing the fields are listed in Table 1. For this analysis, the raw data were reduced using a least-squares self-calibration method²¹; the processing is described in more detail in Supplementary Information.

The random noise level of the maps was computed from two subsets (A, B) containing the odd- and even-numbered frames, respectively, of the observing sequence. The A and B subsets were observed nearly simultaneously and with similar dither patterns and exposures. The difference between maps generated from the A and B subsets should eliminate true celestial sources and stable instrumental effects, and reflect only the random noise of the observations.

Analysis of the background fluctuations must be preceded by steps that eliminate the foreground Galactic stars and the galaxies bright enough to be individually resolved. The primary means of removing these sources is an iterative clipping algorithm which zeroes all pixels (and a fixed number of neighbouring pixels, $N_{\text{mask}} \times N_{\text{mask}}$) that are more than a chosen factor, N_{cut} , above the 1σ r.m.s. variation in the clipped surface brightness. This must be restricted to relatively high

¹Observational Cosmology Laboratory, ²SSAI, ³NASA, Code 665, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Table 1 | Details of analysed fields

Region	QSO 1700 (Ch. 1-4)	HZF (Ch. 1-3)	HZF (Ch. 4)	EGS (Ch. 1-4)
(α, δ)	(255.3, 64.2)	(136.0, 11.6)	(285.7, -17.6)	(215.5, 53.3)
$(l, b)_{\text{Gal}}$	(94.4, 36.1)	(217.5, 34.6)	(18.4, -10.4)	(96.5, 58.9)
$(\lambda, \beta)_{\text{Ecl}}$	(194.3, 83.5)	(135.0, -4.9)	(285.0, 5.0)	(179.9, 60.9)
$\langle t_{\text{obs}} \rangle$ (h)	7.8, 7.8, 7.8, 9.2	0.5, 0.5, 0.5	0.7	1.4, 1.4, 1.4, 1.4
$m_{\text{Vega,lim}}$	22.5, 20.5, 18.25, 17.5	21.5, 19.5, 17	14.5	22.5, 20.75, 18.5, 17.75
Pixel scale (")	0.6	1.2	1.2	1.2
Field size (pix)	1,152 × 512	576 × 256	576 × 256	640 × 384

QSO 1700 was observed during In-Orbit Checkout (IOC) with eight AORs (Astronomical Observation Requests), which used various dither patterns and 200-s frame times, except for two which used 100-s frame times (AOR ID numbers = 7127552, 7127808, 7128064, 7128320, 7128576, 7475968, 7476224, 7476480). Because of the focal plane offset between the shared 3.6/5.8 μm and 4.5/8 μm fields of view, each of these pairs observes separate fields that have a common overlap of $\sim 5' \times 5'$ at all four wavelengths. To provide contrast for the self-calibration algorithm, data from a high-zodiacal light brightness field (HZF) was co-processed for each channel. For 3.6, 4.5 and 5.8 μm , the nearest suitable (200-s frame time) data were observed later during IOC (AOR ID number = 8080896). For the 8- μm data, because nominal 100- and 200-s frame times are split into pairs and quartets of 50-s frames, more nearly contemporaneous observations earlier during IOC were used (AOR ID number = 6849280). For this work we only self-calibrated the six 200-s AORs for 3.6–5.8 μm , but at 8 μm all eight AORs were used as all produce data with 50-s frame times. Observations of the extended Groth Strip area (EGS) provide an additional deep data set for verification of our results. Separate exposure times and limiting magnitudes apply for channels 1–4 respectively. (α, δ) are right ascension and declination, (l, b) and (λ, β) are Galactic and Ecliptic (longitude, latitude) respectively. All coordinates are in degrees. $\langle t_{\text{obs}} \rangle$ is the mean integration time in hours. The maps for the main QSO 1700 field are shown in Supplementary Information.

N_{cut} to leave enough area for a robust Fourier analysis of the map, and to avoid clipping into the background fluctuation distribution. This means that faint sources, the faint outer portions of resolved galaxies, and the faint wings of the point source response function around bright stars cannot be clipped adequately. To remove these low surface brightness sources, we used a CLEAN algorithm²² to model the entire field in each channel. This model, convolved with the full IRAC point spread function, was subtracted from the unclipped regions of the map. Supplementary Information provides details on this process and illustrations of the clipped and model-subtracted images. The final step is the fitting and removal of the zeroth- and first-order components of the background in the

unclipped regions. This is done to minimize power spectrum artefacts due to the clipping, and because the lack of an absolute flux reference measurement and observing constraints prohibit unambiguous determination of these components. With this subtraction, the images represent the fluctuation fields, $\delta F(\mathbf{x})$, at position $\mathbf{x}$ rather than the absolute intensity F . For each observed field, these steps were carried out for images derived from the full data set and from the A and B subsets.

Power spectrum computation and analysis

For each channel, we calculate the power spectra of the fluctuations as a quantitative means of characterizing their scale and amplitude.

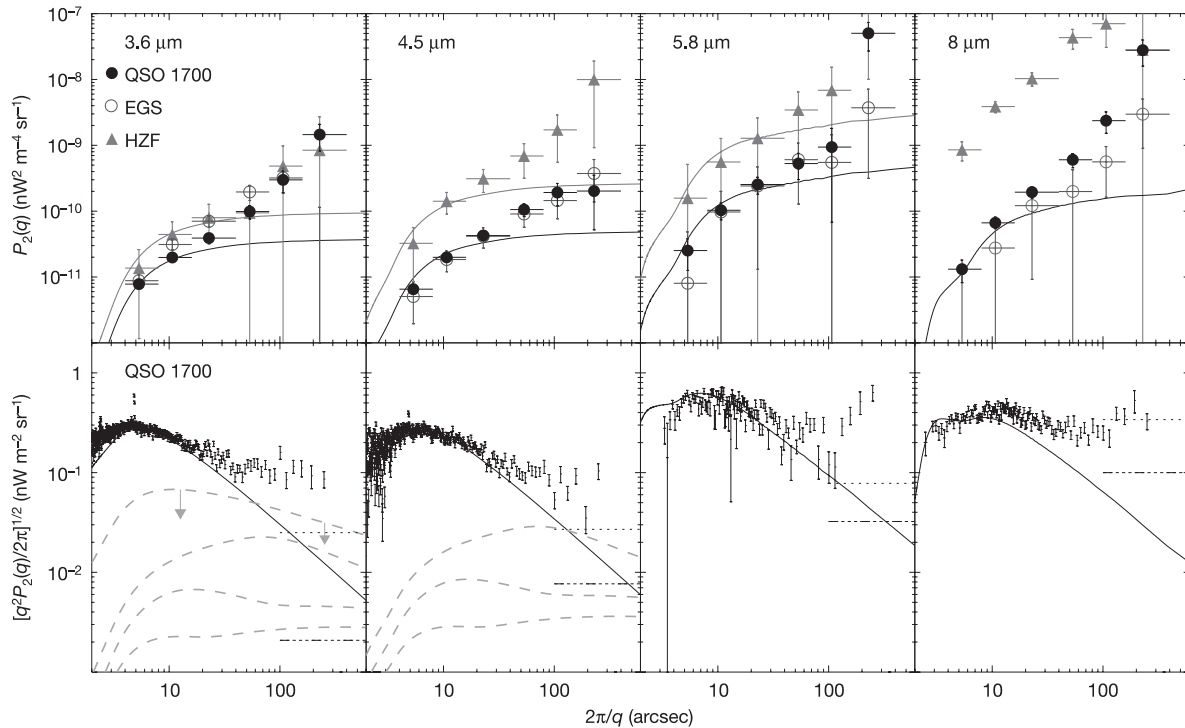


Figure 1 | Spectra of CIB fluctuations. Top, power spectra of signal minus noise from Supplementary Fig. 2 averaged over wide bins to increase signal-to-noise ratio. The errors are $N_q^{-1/2}$, corresponding to the cosmic variance, where N_q is the number of Fourier elements at the given q -bin. Solid lines show the shot noise from remaining galaxies fainter than the limiting magnitude in Table 1. Filled circles and the darkest shade error bars correspond to the QSO 1700 data, open circles and intermediate shade error bars to the EGS data, and triangles with the lightest shade error bars and lines correspond to HZF data. Bottom panels, fluctuations, $[q^2 P_2(q) / 2\pi]^{1/2}$, versus $2\pi/q$ (see text) for the QSO 1700 data. Dashed lines estimate the

contribution from ordinary galaxy populations and ΛCDM density field with $\Delta t = 5$ Gyr: the top dashed line shows the upper limit, which assumes that their (high- z) clustering pattern remained identical to that today at $z = 0$ and the other dashed lines correspond to $\langle z \rangle = 1, 3, 5$ from top to bottom. Solid lines show shot noise from remaining ordinary galaxies. Dotted and dash-dot-dotted lines show the estimated Galactic cirrus and zodiacal light contributions, respectively, assuming the power spectra to be $P(q) \propto q^n$ with $n = -2$, typical of cloud distributions. The observed cirrus power spectrum is a little steeper with $n \approx -2.5$ to -3 (refs 14, 37, 38, 39) in which case the lines will have a slope of $(3 + n)/2 = 0.25$ to 0.5 .

Table 2 | Cross-correlation coefficients of fluctuations

Channels	$N_{\text{cut}} = 4$	$N_{\text{cut}} = 2$
1: 2	52	12
1: 3	7	0.6
1: 4	10	4

The cross-correlation coefficients are shown for various clipping thresholds and are in units of that of random sample $\mathcal{R}/\sigma_{\mathcal{R}}$. These correlations were evaluated for an evenly covered region of 512^2 pixels common to all four channels for the QSO 1700 field. For random uncorrelated samples, δ_1, δ_2 , of N_{pixels} , the correlation coefficient, $\mathcal{R} \equiv \langle \delta_1 \delta_2 \rangle / (\langle \delta_1^2 \rangle \langle \delta_2^2 \rangle)^{1/2}$, should be zero with dispersion $\sigma_{\mathcal{R}} = N_{\text{pixels}}^{-1/2}$. In fact, for $N_{\text{mask}} = 3$, we find that down to even $N_{\text{cut}} = 2$, when only 6% of the pixels remain, the correlations between the channels remain statistically significant. Simple simulations containing the appropriate levels of the instrument noise and a power law component of the CIB gave somewhat larger mean values of $\mathcal{R}$ with the measured values lying within a 95% confidence level; incorporating the possibility that not all of the remaining ordinary galaxies are the same at each wavelength would, however, reduce the mean values of $\mathcal{R}$.

The power that remains after subtraction of the random noise component is insensitive to the details of the source clipping, and is statistically correlated between channels. This confirms that the fluctuation signal does have a celestial origin. The shape and amplitude of the power spectra are not consistent with significant contributions from the cirrus of the interstellar medium (except perhaps at $8 \mu\text{m}$) or from the zodiacal light from local interplanetary dust.

The fluctuation field, $\delta F(\mathbf{x})$, was weighted by the observation time t_{obs} in each pixel, $w(\mathbf{x}) \propto t_{\text{obs}}(\mathbf{x})$, and its Fourier transform, $f(\mathbf{q}) = \int \delta F(\mathbf{x}) w(\mathbf{x}) \exp(-i\mathbf{x} \cdot \mathbf{q}) d^2x$ as function of the angular wave-number $\mathbf{q}$ calculated using the fast Fourier transform. (The weighting is necessary to minimize the noise variations across the image, but we also performed the same analysis without it and verified that the weight adds no structure to the resultant power spectrum. This is because the weights are relatively flat across the image.) The power spectrum is $P_2(q) = \langle |f(\mathbf{q})|^2 \rangle$, with the average taken over all the Fourier elements N_q corresponding to the given q . A typical flux fluctuation is $\sqrt{q^2 P_2(q)/2\pi}$ on the angular scale of wavelength $2\pi/q$. In the Supplementary Information we show the final power spectrum of the diffuse flux fluctuations, $P_S(q)$, and the noise, $P_N(q)$, of the data sets. We find significant excess of the large-scale fluctuations over the instrument noise.

Figure 1 shows the excess power spectrum, $P_S(q) - P_N(q)$, of the diffuse light after the instrument noise has been subtracted at 3.6, 4.5, 5.8 and $8 \mu\text{m}$. The errors correspond to standard deviation for the cosmic variance, i.e. the relative error on each $P_S(q)$, $P_N(q)$ is $N_q^{-1/2}$. There is a clear positive residual whose power spectrum is significantly different from white noise and has substantial correlations all the way to the largest scales probed. Possible sources of these large-scale correlations can be artefacts of the analysis procedure (such as clipping and fast Fourier transforms), instrumental artefacts, local Solar System or Galactic emission, or relatively nearby extragalactic sources and/or more distant cosmological sources. In what follows we discuss their relative contributions.

Residual emission from the wings of incompletely clipped sources can give rise to spurious fluctuations, but the power spectrum of these fluctuations should depend on the clipping parameters N_{cut} and N_{mask} . We tested the contributions from these residuals in various ways. For a given N_{cut} we varied N_{mask} from 3 to 7, significantly reducing any residual wings and increasing the fraction of the clipped pixels, but found negligible (less than a few per cent) variations in the final $P_2(q)$. We also clipped down to progressively lower values of N_{cut} . For $N_{\text{cut}} \lesssim 3.5$, too few pixels remain for robust Fourier analysis, so in these cases we computed $C(\theta)$, the correlation function of the diffuse emission, as a function of angular separation θ , related to the power spectrum by a one-dimensional Legendre transformation. It is consistent with the fluctuations in Fig. 1 as discussed and shown in Supplementary Information.

Instrument noise contributions to the fluctuations were evaluated as the power spectra of $\frac{1}{2}(\delta F_A - \delta F_B)$ using the A–B subset images.

As shown in the Supplementary Information, random instrument noise has an approximately white spectrum. The power spectra of the final data sets have a much larger amplitude than the noise (especially in channels 1 and 2) until the convolution with the beam tapers off the signals from the sky at the smallest angular scales. The instrument noise spectra are unaffected by the beam and are uncorrelated from channel to channel, as expected. However, the full data set fluctuation fields show statistically significant correlations between channels, as shown in Table 2. This means that we see the same fluctuation field in addition to (different) noise in all four channels. In Supplementary Information we describe tests to assess the contributions of possible instrumental systematic errors. They indicate that systematic instrumental effects are unlikely to lead to the signal shown in Fig. 1.

The best assessment of zodiacal light contributions to the power spectrum comes from the examination of EGS observations taken at two epochs six months apart. Because any anisotropies in the zodiacal light cloud will not remain fixed in celestial coordinates over this interval, the difference in the fluctuation fields at these two epochs should eliminate Galactic and extragalactic signals and yield a power spectrum of the zodiacal light fluctuations added to the instrument noise and possible systematic errors. The fluctuation levels of these difference maps set an upper limit on the zodiacal light contribution of $<0.1 \text{ nW m}^{-2} \text{ sr}^{-1}$ at $8 \mu\text{m}$. Scaling this result to the other channels by interpolating the observed zodiacal light spectrum²³ leads to zodiacal light contributions to the fluctuations that are comfortably below the detections in Fig. 1 in the other channels as well.

Our assessment of the contribution of the infrared cirrus (that is, interstellar clouds of neutral gas and dust) to the power spectra is derived from the $8 \mu\text{m}$ HZF field, which lies at low Galactic latitude and is visibly contaminated by cirrus. Assuming that the large-scale fluctuations in this field are due to the cirrus, relative fluctuations of the $8 \mu\text{m}$ cirrus are $\sim 1\%$ of the mean cirrus flux level. A similar level of relative cirrus fluctuations in the QSO 1700 field would have an amplitude of $\sim 0.3 \text{ nW m}^{-2} \text{ sr}^{-1}$. This is not significantly lower than the $8 \mu\text{m}$ amplitude observed in Fig. 1. Therefore, we cannot at present eliminate the possibility that the fluctuations at $8 \mu\text{m}$ are dominated by cirrus. However, the spectrum of the interstellar medium emission should drop sharply at shorter wavelengths as the PAH (polycyclic aromatic hydrocarbons) emission bands that dominate at $8 \mu\text{m}$ become less significant. Given this estimate of the cirrus contribution at $8 \mu\text{m}$ we estimate the amplitude of large-scale fluctuations for channels 1–3 as $\sim 0.03, 0.03$ and $0.08 \text{ nW m}^{-2} \text{ sr}^{-1}$,

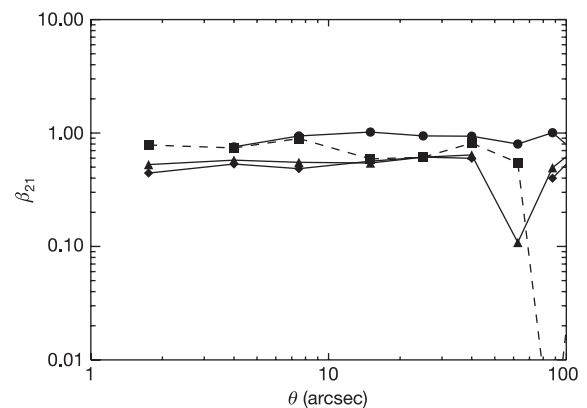


Figure 2 | Colour properties of clipped maps. Estimates of colour between channels 1 and 2. Circles correspond to $\sqrt{P_2/P_1}$, averaged over the bins centred at these angles, diamonds to $\langle \delta F_2(x) \delta F_1(x + \theta) \rangle / \langle \delta F_1(x) \delta F_1(x + \theta) \rangle$ for $N_{\text{cut}} = 4$, and triangles to the same quantity averaged over maps with $N_{\text{cut}} = 4, 3, 2.5, 2$. Squares correspond to β_{41}/β_{42} averaged over maps with $N_{\text{cut}} = 4, 3, 2.5, 2$. (See text for definitions of variables.)

which are well below the observed fluctuations in these channels.

Colour is another important criterion for testing the origins of the signal in Fig. 1. The cross-correlations between the channels for the common area of the QSO 1700 field are statistically significant and they also strengthen at the larger scales where the noise contribution is smaller, as verified by smoothing the maps. These significant cross-correlations allow us to examine the corresponding colours of the fluctuations. We made several estimates of the colours, β_{n1} , between channels n and 1: first, as the square root of ratio of the power spectra, and second, as $\beta_{n1}(\theta) = \langle \delta F_n(x) \delta F_1(x + \theta) \rangle / \langle \delta F_1(x) \delta F_1(x + \theta) \rangle$ evaluated over the 512^2 pixel field common to all four channels, where a further consistency check comes from comparison between β_{21} and β_{41}/β_{42} . For channels 1 and 2 these estimates are shown in Fig. 2, and appear roughly independent of angular scale and mutually consistent. The instrument noise is too large to enable robust estimates involving channel 3 ($5.8 \mu\text{m}$). These derived colours indicate that the fluctuation signal in Fig. 1 has an energy distribution that is approximately flat to slowly rising with wavelength in νI_ν , where I_ν is intensity at frequency ν . The energy spectrum rules out contributions from remaining Galactic stars, but probably cannot be used to distinguish between ordinary galaxies and population III objects without additional detailed modelling of both sets of sources.

CIB fluctuations from extragalactic sources

There are two extragalactic classes of contributors to CIB fluctuations: ‘ordinary’ galaxies containing normal stellar populations I and II, and the objects that preceded them, population III stars. The square of the CIB fluctuation in band ν produced by cosmological sources that existed over time period Δt is given by the power-spectrum version of the Limber equation (see ref. 2 and Supplementary Information):

$$\frac{q^2 P_2(q)}{2\pi} = \Delta t \int \left(\frac{dI_{\nu'}}{dt} \right)^2 \Delta^2(q d_A^{-1}; z) dt \quad (1)$$

where $\nu' = \nu(1+z)$ is the rest frequency of the emitters, d_A is the comoving angular diameter distance, and

$$\Delta(k) = \sqrt{\frac{1}{2\pi} \frac{k^2 P_3(k)}{c \Delta t}} \quad (2)$$

is the fluctuation in the number of sources within a volume $k^{-2} c \Delta t$ and k is the spatial wavenumber. The fluctuation of the CIB with mean flux F_{CIB} on angular scale $\sim 2\pi/q$ can then be expressed as $\delta F_{\text{CIB}} = F_{\text{CIB}} \Delta(q d_A^{-1}(\langle z \rangle))$, where $\langle z \rangle$ is the suitably averaged effective redshift. The relative CIB fluctuation is the suitably averaged fluctuation in the source counts over a cylinder of radius $q^{-1} d_A(\langle z \rangle)$ and length $c \Delta t$. Three things would lead to larger fluctuations; (1) a population that, after removing constituents brighter than some limit, leaves a substantial mean CIB flux (increase F_{CIB}), (2) populations that existed for a shorter time (increase $\langle \Delta \rangle$ by decreasing Δt), and (3) populations that formed out of rare peaks of the underlying density field leading to biased and significantly amplified²⁴, clustering properties (increase $\langle \Delta \rangle$ by increasing $P_3(k)$).

CIB fluctuations from remaining ordinary galaxies

The fluctuations produced by ordinary galaxies contain two components: first, shot noise from discrete galaxies, and second, galaxy clustering from the primeval density field. The amplitude of the first component can be estimated directly from galaxy counts. In order to estimate the contribution from the second component we proceed as follows: from galaxy count data we estimate the total CIB flux produced by the remaining galaxies fainter than our clipping threshold. Ordinary galaxies occupy an era of $\Delta t \gtrsim$ a few Gyr. Their present-day clustering pattern on the relevant scales is well measured today. The clustering pattern evolution can be extrapolated to earlier times assuming the ‘concordance’ Λ CDM model (see below). These parameters (flux, Δt , clustering pattern and its evolution) then allow us to estimate the contribution to the CIB fluctuations via equations (1) and (2). Because galaxy clustering is weaker at earlier times, an upper limit on the contribution from ordinary galaxies is obtained assuming that the clustering at early times remained the same as at $z = 0$.

The shot noise component contributed by the ordinary galaxies to the CIB angular spectrum was estimated directly from galaxy count data: each magnitude bin Δm with galaxies of flux $f(m)$ would contribute a power spectrum of $P_{\text{sn}} = f^2(m) \frac{dN_{\text{gal}}}{dm} \Delta m$. Figure 1 shows the shot noise component convolved with the IRAC beam for the galaxies at the limiting magnitude in Table 1. The limiting magnitudes are qualitative estimates of where the SExtractor²⁵ number counts (in 0.5 magnitude bins) begin to drop sharply due

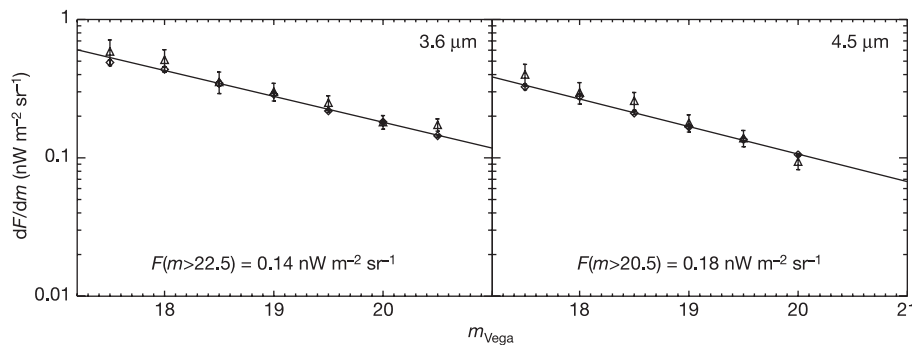


Figure 3 | Contribution to CIB flux from Spitzer IRAC galaxy counts at 3.6 μm and 4.5 μm . Vertical axis shows the flux, F , produced by galaxies in the magnitude range $[m, m + dm]$ shown on the horizontal axis. Errors correspond to standard deviation in poissonian number counts. The value of dF/dm is evaluated from counts data and is a rapidly decreasing function of m of the form $dF/dm \propto \exp(bm)$ with $b \approx 0.4$ (the least squares fits are shown with solid lines). (At the other two IRAC channels the uncertainty in the fall-off is much greater, but we show below that definite conclusions can already be reached from the data at 3.6 and 4.5 μm). Assuming that the functional form of dF/dm does not begin to rise appreciably at still fainter magnitudes gives the CIB contribution from galaxies fainter than m_0 of $F(m > m_0) = b^{-1} dF/dm|_{m_0}$. For the QSO 1700 field, we eliminate galaxies brighter than $m_{\text{Vega}} \approx 22.5$ at 3.6 μm and $m_{\text{Vega}} \approx 20.5$ at 4.5 μm corresponding to AB magnitudes ~ 25 and 24, respectively. The annotation

in each panel shows numbers that correspond to the extrapolated total flux from galaxies fainter than the magnitude limits in Table 1 using the least-squares fits shown with solid lines. One can expect that galactic spectra at the appropriate range of wavelengths are at most as steep as that of Vega, a star with a Rayleigh-Jeans black body spectral fall-off at these wavelengths. If so, they would generally have $K - m_{3.6} > 0$, leaving galaxies with $K \gtrsim 22$ (see, for example, ref. 40). Galaxies at $K > 20$ have median redshift $\gtrsim 1$ (ref. 41), so the above argument would place the galaxies remaining at 4.5 μm at $z \gtrsim 1$ and those remaining at 3.6 μm still farther out to higher z . Comparison with the Lyman-break galaxy candidates in the same area shows that a substantial fraction of these galaxies are at $z \approx 2-4$, with $z \sim 3$ being the median redshift²⁰. Similarly, a substantial part of these galaxies are confirmed to be star-forming systems at $z \approx 2.3$ (ref. 42).

to incompleteness. The shot noise fits the observed diffuse light fluctuations well at small angles, but the shot-noise contribution from ordinary galaxies cannot make a substantial contribution to the large-scale power of the diffuse light in Fig. 1.

We assume the 'concordance' Λ CDM model Universe with flat geometry (total density parameter $\Omega_{\text{total}} = 1$) dominated by a cosmological constant, $\Omega_{\Lambda} \approx 0.7$, with the rest coming from cold dark matter and ordinary baryons in proportions suggested by measurements of the CMB anisotropies²⁶ and high- z supernovae²⁷. With $P_3(k)$ taken from the Λ CDM concordance model, the power in fluctuations from the clustering of ordinary galaxies depends on the net flux produced by the remaining ordinary galaxies (via F_{CIB}) and their typical redshifts (via $\Delta(qd_A^{-1})$). The total flux from galaxies fainter than the limiting magnitude in Table 1 was estimated from the Spitzer IRAC galaxy counts¹⁸, and is shown in Fig. 3. The flux contributed to the CIB from the remaining galaxies is between 0.1 and 0.2 $\text{nW m}^{-2} \text{sr}^{-1}$ and this amplitude is much less than the excess CIB at these wavelengths² indicating that the excess CIB flux is produced by populations that are still farther out than the ordinary galaxies remaining in the data. This (low) value of the remaining flux can also be derived from the small amplitude of the residual shot-noise contribution to the fluctuations in the confusion-limited data sets. Thus, in order to explain the amplitudes shown in Fig. 1, the remaining ordinary galaxy populations would need to produce relative flux fluctuations of the order of $\geq 100\%$ from their clustering. Without follow-up spectroscopy it is difficult to determine with high precision the range of redshifts of the remaining ordinary galaxies, but approximate estimates can be made, and are summarized in the Fig. 3 legend to be $z \geq 1$.

In flat cosmology with $\Omega_{\Lambda} = 0.7$, one arcminute subtends a comoving scale between 0.7 and 1.6 h^{-1} Mpc at z between 1 and 5. The present-day three-dimensional power spectrum of galaxy clustering, $P_{3,\text{gal}}(k)$, is described well by the concordance Λ CDM model^{28,29} and we expect that on arcminute scales the density field was in the linear to quasi-linear regime at the redshifts probed by the remaining galaxies between 3.6 and 8 μm . At smaller scales, nonlinear corrections to evolution were computed following ref. 30. The resultant CIB fluctuation from the remaining ordinary galaxies, producing mean CIB $F_{\text{CIB,og}} = 0.14 \text{ nW m}^{-2} \text{sr}^{-1}$, times $\Delta(\propto(\Delta t)^{-1/2})$ is shown in Fig. 1 for $\langle z \rangle = 1, 3, 5$ and $\Delta t = 5$ Gyr, corresponding to the age of the Universe at $z \approx 1$. For the Λ CDM model, the relative fluctuations in the CIB on arcminute scales would be of the order of $\langle \Delta \rangle \approx (2-10) \times 10^{-2} (\Delta t)^{-0.5}$ from galaxies with $\langle z \rangle = 1-5$ assuming no biasing (here Δt is in units of 5 Gyr). Combining this with the above values for the diffuse flux from the remaining ordinary galaxies would lead to $\delta F \lesssim (1-2) \times 10^{-2} \text{ nW m}^{-2} \text{sr}^{-1}$ in all the channels. While biasing may increase the relative fluctuations, with reasonable bias factors for galaxies lying at $z \approx$ (a few), the diffuse light fluctuations would still be very small compared to those in Fig. 1 and are unlikely to account for fluctuations of amplitude $\sim (0.1-0.5) \text{ nW m}^{-2} \text{sr}^{-1}$ at arcminute scales. An upper limit on the CIB fluctuations can be evaluated assuming the same clustering pattern for the remaining galaxies as at the present epoch, $z = 0$, that is, that their two-point correlation function is given by $\xi = (r/r_*)^{-1.7}$ with $r_* = 5.5 h^{-1}$ Mpc (ref. 31); its Δ times $F_{\text{CIB,og}}$ is also shown in Fig. 1 and is much below the signal we measure. Conversely, one can evaluate the clustering strength needed to account for the observed fluctuations: a 100% relative fluctuation from galaxies at $z \geq 1$ clustered with $\xi = (r/r_*)^{-1.7}$ would require $r_* \geq 25 h^{-1}$ Mpc. This corresponds to the effective bias factor ≥ 5 , which is significantly higher than expected from gravitational clustering evolution in the Λ CDM universe³². In fact, direct measurements of clustering for relatively nearby IRAC galaxies with flux $> 32 \mu\text{Jy}$ at 3.6 μm (~ 5 mag brighter than the limit reached for remaining galaxies in the QSO 1700 field)³³ find the projected two-point correlation amplitude on $\sim$ arcminute scales of $w_{>32\mu\text{Jy}}(\theta \geq 1') < 4 \times 10^{-2}$ corresponding to relative fluctuations in CIB from these galaxies of amplitude

$\sim \sqrt{w} \lesssim 0.2$. Fainter galaxies are expected to have an even lower correlation amplitude.

The QSO 1700 field seems to contain an overdensity of Lyman-break galaxies at $z \approx 2.3$ (ref. 34) which could lead to a larger Δ and CIB fluctuations for this region. However, we see similar levels of fluctuations for the other fields located at very different parts of the sky, making it unlikely that the overdensity claimed in ref. 34 can account for our signal. When account is made of the different shot noise levels from the remaining galaxies, the fluctuations seen in the different fields have consistent power spectra within the statistical uncertainties. (An exception is channel 4 HZF data, located at low Galactic Latitude, and clearly dominated by Galactic cirrus.)

CIB fluctuations from the population III era

Population III objects at $z \approx 10-30$ are expected to precede ordinary galaxy populations. One can expect on fairly general grounds⁸ that, if massive, they would contribute significantly to the NIR CIB, in terms of both its mean level and anisotropies. Intuitive reasons are discussed in ref. 2, but as equations (1) and (2) show, they are mostly related to: (1) if massive, population III were very efficient light emitters, (2) their era probably lasted a shorter time, Δt , than that of the ordinary galaxies, leading to larger Δ in equation (2), and (3) they should have formed out of high peaks of the density field whose correlation function is strongly amplified. The NIR also probes ultraviolet to visible parts of the electromagnetic spectrum at $z \approx 10-30$, where most of their emission is produced¹².

The amplitude of the CIB anisotropies remaining in the present data implies that the remaining CIB originates from still fainter objects. Can the observed amplitudes of fluctuations in Fig. 1 be accounted for by energetic sources at high z ? Population III stars can produce significant NIR CIB levels¹², $\geq 1 \text{ nW m}^{-2} \text{sr}^{-1}$ and, for example, the NIR CIB excess over that from ordinary galaxies at 3.6 μm is $8.7 \pm 3.1 \text{ nW m}^{-2} \text{sr}^{-1}$ (ref. 2). At other NIR wavelengths $> 1 \mu\text{m}$, the mean NIR CIB excess is also more than a few $\text{nW m}^{-2} \text{sr}^{-1}$ (refs 15, 43-49). Thus population III objects would require smaller relative CIB fluctuations. Because individual population III systems are small, yet numerous, the shot-noise component of the CIB from them is small and, in any case, is already absorbed in the shot-noise shown in Fig. 1. Additionally, their $\langle \Delta \rangle$ would be amplified by the much shorter Δt and (significant) biasing, and population III contribution would dominate the diffuse light fluctuations in Fig. 1. Detailed theoretical interpretation of the results in terms of the population III era models exceeds the scope of this Article, but qualitative comparison can be made by estimating the typical value of the relative CIB fluctuation, Δ , corresponding to that era. The fraction of the population III haloes was calculated assuming that they form from the Λ CDM density field in haloes where the virial temperature $T_{\text{vir}} \geq 2,000$ K to enable efficient molecular hydrogen cooling³⁵. Biasing was treated using the gravitational clustering prescription from ref. 36 in the Λ CDM model. (Nonlinear evolution effects are small on the angular scales and redshifts probed here.) Assuming $\Delta t \approx 300$ Myr, which corresponds to the age of the Universe at $z = 0$, we get typical values of $\Delta \approx 0.1-0.2$ at z between 10 and 20. This order-of-magnitude evaluation shows that the levels of $\sim 0.1-0.3 \text{ nW m}^{-2} \text{sr}^{-1}$ at 3.6 to 8 μm on arcminute scales can be accounted for by population III emissions if their total flux contribution is $> 1 \text{ nW m}^{-2} \text{sr}^{-1}$, which is reasonable, as discussed earlier.

Earlier studies of CIB fluctuations¹⁴⁻¹⁶ contained significant contributions from relatively bright galaxies, making it difficult to isolate the possible CIB fluctuations from the very early times. The contribution to CIB fluctuations from remaining galaxies is a function of the limiting magnitude below which galaxies are removed. With the Spitzer IRAC data we could identify and remove galaxies to very faint limits of flux $\geq 0.3 \mu\text{Jy}$. This limit is, at last, sufficiently low to push the residual contribution from ordinary galaxies along the line-of-sight below the level of the excess signal at larger angular scales. If our

interpretation is correct and the signal we detect comes from population III located at much higher z , the amplitude of the CIB fluctuations on scales where galaxy shot-noise is negligible should remain the same as fainter ordinary galaxies are removed with deeper clipping. This is true as far as we can test with this data, and would certainly be verifiable with longer-exposure data.

At these z the IRAC bands probe the rest-frame wavelengths between 0.2 and 0.8 μm , where the energy spectra of individual population III objects is dominated by free-free emission and is a slowly rising function of wavelength¹²; this would be consistent with the spectrum of the CIB anisotropies in Fig. 1. NIR observations at shorter wavelengths would be particularly important in confirming the redshifts where the CIB fluctuations originate, as there should be a significant drop in the power of the fluctuations at the rest-frame Lyman limit wavelength.

Received 3 May; accepted 12 August 2005.

- Hauser, M. G. & Dwek, E. The cosmic infrared background: Measurements and implications. *Annu. Rev. Astron. Astrophys.* **39**, 249–307 (2001).
- Kashlinsky, A. Cosmic infrared background and early galaxy evolution. *Phys. Rep.* **409**, 361–438 (2005).
- Kogut, A. *et al.* First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Temperature-polarization correlation. *Astrophys. J. Suppl.* **148**, 161–173 (2003).
- Abel, T. *et al.* The formation of the first star in the Universe. *Science* **295**, 93–98 (2002).
- Bromm, V. *et al.* Forming the first stars in the Universe: The fragmentation of primordial gas. *Astrophys. J.* **527**, L5–L8 (1999).
- Bromm, V. & Larson, R. B. The first stars. *Annu. Rev. Astron. Astrophys.* **42**, 79–118 (2004).
- Rees, M. J. Origin of pregalactic microwave background. *Nature* **275**, 35–37 (1978).
- Kashlinsky, A., Arendt, R., Gardner, J. P., Mather, J. & Moseley, S. H. Detecting population III stars through observations of near-infrared cosmic infrared background anisotropies. *Astrophys. J.* **608**, 1–9 (2004).
- Kashlinsky, A., Mather, J. & Odenwald, S. Clustering of the diffuse infrared light from the COBE DIRBE maps. IV. Overall results and possible interpretations. Preprint (1999).
- Salvaterra, R. & Ferrara, A. The imprint of the cosmic dark ages on the near-infrared background. *Mon. Not. R. Astron. Soc.* **339**, 973–982 (2003).
- Magliocchetti, M., Salvaterra, R. & Ferrara, A. First stars contribution to the near-infrared background fluctuations. *Mon. Not. R. Astron. Soc.* **342**, L25–L29 (2003).
- Santos, M., Bromm, V. & Kamionkowski, M. The contribution of the first stars to the cosmic infrared background. *Mon. Not. R. Astron. Soc.* **336**, 1082–1092 (2002).
- Cooray, A., Bock, J., Keating, B., Lange, A. & Matsumoto, T. First star signature in infrared background anisotropies. *Astrophys. J.* **606**, 611–624 (2004).
- Kashlinsky, A. & Odenwald, S. Clustering of the diffuse infrared light from the COBE DIRBE maps. III. Power spectrum analysis and excess isotropic component of fluctuations. *Astrophys. J.* **528**, 74–95 (2000).
- Matsumoto, T. *et al.* Infrared telescope in space observations of the near-infrared extra-galactic background light. *Astrophys. J.* **626**, 31–43 (2005).
- Kashlinsky, A., Odenwald, S., Mather, J., Skrutskie, M. & Cutri, R. Detection of small-scale fluctuations in the near-infrared cosmic infrared background from long-exposure 2MASS fields. *Astrophys. J.* **579**, L53–L57 (2002).
- Odenwald, S., Kashlinsky, A., Mather, J., Skrutskie, M. & Cutri, R. Analysis of the diffuse near-infrared emission from Two-Micron All-Sky Survey deep integration data: Foregrounds versus the cosmic infrared background. *Astrophys. J.* **583**, 535–550 (2003).
- Fazio, G. G. *et al.* Number counts at $3\mu\text{m} < \lambda < 10\mu\text{m}$ from the Spitzer Space Telescope. *Astrophys. J. Suppl.* **154**, 39–43 (2004).
- Fazio, G. G. *et al.* The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J. Suppl.* **154**, 10–17 (2004).
- Barmby, P. *et al.* Deep mid-infrared observations of Lyman break galaxies. *Astrophys. J. Suppl.* **154**, 97–102 (2004).
- Fixsen, D. J., Moseley, S. H. & Arendt, R. G. Calibrating array detectors. *Astrophys. J. Suppl.* **128**, 651–658 (2000).
- Högbom, J. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astrophys. J. Suppl.* **15**, 417–426 (1974).
- Kelsall, T. *et al.* The COBE Diffuse Infrared Background Experiment search for the cosmic infrared background. II. Model of the interplanetary dust cloud. *Astrophys. J.* **508**, 44–73 (1998).
- Kaiser, N. On the spatial correlations of Abell clusters. *Astrophys. J.* **284**, L9–L12 (1984).
- Bertin, E. & Arnouts, S. SExtractor: Software for source extraction. *Astron. Astrophys. Suppl.* **117**, 393–404 (1996).
- Bennett, C. *et al.* First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. *Astrophys. J. Suppl.* **148**, 1–37, (2003).
- Perlmutter, S. *et al.* Measurements of omega and lambda from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
- Efstathiou, G., Sutherland, W. J. & Maddox, S. J. The cosmological constant and cold dark matter. *Nature* **348**, 705–707 (1990).
- Tegmark, M. *et al.* Cosmological parameters from SDSS and WMAP. *Phys. Rev. D* **69**, 103501–103527 (2004).
- Peacock, J. & Dodds, S. J. Non-linear evolution of cosmological power spectra, 1996. *Mon. Not. R. Astron. Soc.* **280**, L19–L26 (1996).
- Maddox, S., Efstathiou, G., Sutherland, W. & Loveday, J. Galaxy correlations on large scales. *Mon. Not. R. Astron. Soc.* **242**, 43P–47P (1990).
- Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).
- Oliver, S. *et al.* Angular clustering of galaxies at 3.6 microns from the Spitzer Wide-area Infrared Extragalactic (SWIRE) survey. *Astrophys. J. Suppl.* **154**, 30–34 (2004).
- Steidel, C. *et al.* Spectroscopic identification of a protocluster at $z = 2.300$: Environmental dependence of galaxy properties at high redshift. *Astrophys. J.* **626**, 44–50 (2005).
- Miralda-Escude, J. The dark age of the Universe. *Science* **300**, 1904–1909 (2003).
- Kashlinsky, A. Reconstructing the spectrum of the pregalactic density field from astronomical data. *Astrophys. J.* **492**, 1–28 (1998).
- Gautier, T. N., Boulanger, F., Perault, M. & Puget, J. L. A calculation of confusion noise due to infrared cirrus. *Astron. J.* **103**, 1313–1324 (1992).
- Ingalls, J. G. *et al.* Structure and colors of diffuse emission in the Spitzer galactic first look survey. *Astrophys. J.* **154**, 281–285 (2004).
- Wright, E. L. Angular power spectra of the COBE DIRBE maps. *Astrophys. J.* **496**, 1–8 (1998).
- Eisenhardt, P. R. *et al.* The Infrared Array Camera (IRAC) shallow survey. *Astrophys. J. Suppl.* **154**, 48–53 (2004).
- Cowie, L. L. *et al.* New insight on galaxy formation and evolution from Keck spectroscopy of the Hawaii Deep Fields. *Astron. J.* **112**, 839–864 (1996).
- Shapley, A. E. *et al.* Ultraviolet to mid-infrared observations of star-forming galaxies at $z \sim 2$: Stellar masses and stellar populations. *Astrophys. J.* **626**, 698–722 (2005).
- Hauser, M. G. *et al.* The COBE diffuse infrared background experiment search for the cosmic infrared background. I. Limits and detections. *Astrophys. J.* **508**, 25–43 (1998).
- Wright, E. L. DIRBE minus 2MASS: Confirming the cosmic infrared background at 2.2 microns. *Astrophys. J.* **553**, 538–544 (2001).
- Arendt, R. *et al.* The COBE diffuse infrared background experiment search for the cosmic infrared background. III. Separation of galactic emission from the infrared sky brightness. *Astrophys. J.* **508**, 74–105 (1998).
- Dwek, D. & Arendt, R. A tentative detection of the cosmic infrared background at 3.5 μm from COBE/DIRBE observations. *Astrophys. J.* **508**, L9–L12 (1998).
- Wright, E. L. & Reese, E. D. Detection of the cosmic infrared background at 2.2 and 3.5 microns using DIRBE observations. *Astrophys. J.* **545**, 43–55 (2000).
- Gorjian, V., Wright, E. L. & Chary, R. R. Tentative detection of the cosmic infrared background at 2.2 and 3.5 microns using ground-based and space-based observations. *Astrophys. J.* **536**, 550–560 (2000).
- Cambresy, L., Reach, W. T., Beichman, C. A. & Jarrett, T. H. The cosmic infrared background at 1.25 and 2.2 microns using DIRBE and 2MASS: A contribution not due to galaxies? *Astrophys. J.* **555**, 563–571 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Fazio for access to the IRAC Deep Survey data and D. Fixsen and G. Hinshaw for comments on drafts of this paper. This Article reports work supported by the National Science Foundation, and which is based on observations made with the Spitzer Space Telescope (this telescope is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA). Support for this work was also provided by NASA through an award issued by JPL/Caltech.

Author Contributions A.K. is responsible for the idea, clipping the maps, power spectrum and correlation analyses, evaluating the extragalactic contributions and writing the paper. R.G.A. is responsible for the images for analysis, providing the model of the resolved sources with the IRAC PSF, and evaluating systematics, instrument, and zodiacal and cirrus contributions. J.M. and S.H.M. developed analysis strategy and searched for alternative explanations for the fluctuations. All authors provided critical review of the analysis techniques, results and manuscript.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K. (kashlinsky@stars.gsfc.nasa.gov).

Radiocarbon dating of interstratified Neanderthal and early modern human occupations at the Chatelperronian type-site

Brad Gravina¹, Paul Mellars¹ & Christopher Bronk Ramsey²

The question of the coexistence and potential interaction between the last Neanderthal and the earliest intrusive populations of anatomically modern humans in Europe has recently emerged as a topic of lively debate in the archaeological and anthropological literature. Here we report the results of radiocarbon accelerator dating for what has been reported as an interstratified sequence of late Neanderthal and early anatomically modern occupations at the French type-site of the Chatelperronian, the Grotte des Fées de Châtelperron, in east-central France. The radiocarbon measurements seem to provide the earliest secure dates for the presence of Aurignacian technology—and from this, we infer the presence of anatomically modern human populations—in France.

It is widely believed that the archaeological traces of the last Neanderthal and earliest anatomically modern human populations in western Europe are represented by the Chatelperronian and Aurignacian cultures, respectively^{1–10}. The Chatelperronian culture is confined essentially to western France and northern Spain, and shows strong technological links with the immediately preceding Mousterian technologies. The Aurignacian culture is distributed over the whole of western, central and eastern Europe, and apparently reflects a major population dispersal across these regions (deriving ultimately from the Near Eastern region) over the period approximately 43,000–36,000 before present (BP) (uncalibrated radiocarbon years)⁹. Skeletal remains associated with the Chatelperronian and Aurignacian industries, although not abundant, bear out these correlations^{1,2,9,11,12}.

One inescapable implication of this population dispersal scenario is that there would inevitably have been some period of coexistence and chronological overlap between the final Neanderthal and early anatomically modern populations in western Europe (as in other parts of Europe), with corresponding opportunities for various forms of contact, demographic interaction and potential exchanges of technology or other aspects of culture between the two populations. Hitherto, direct archaeological evidence for this overlap has proved controversial, with the suggestion that supposedly direct ‘interstratifications’ of Chatelperronian and Aurignacian levels at three separate sites in western France and northern Spain (Roc de Combe and Le Piage in southwest France, and El Pendo in northwest Spain) might in fact reflect serious confusions or misinterpretations of the stratigraphy at these sites^{5,6,13,14}. The results reported here seem to provide clear evidence for a direct interstratification of distinctively Chatelperronian and Aurignacian occupations at the type-site of Châtelperron itself, closely dated by a sequence of 13 high-resolution radiocarbon accelerator mass spectrometry (AMS) measurements by the Oxford Radiocarbon Laboratory.

The Châtelperron site

The cave of the Grotte des Fées at Châtelperron lies in the small valley of the Graveron, between the larger valleys of the Loire and the Allier, on the northern flanks of the Massif Central, some 30 km southeast of the town of Moulins (Allier Department). Following its discovery during railway construction in the 1840s, excavations were initially undertaken by A. Poirrier around 1850 and subsequently by G. Bailleau between 1867 and 1872—excavations that removed most of the filling from the central part of the cave. Bailleau recognized only one major archaeological level in his excavations, which provided the material on which Breuil based his original definition of the Chatelperronian industry (initially designated as ‘Aurignacien Inférieur’) in 1910 (ref. 15).

The most recent and important excavations on the site were undertaken by Henri Delporte (subsequently Director of the National Museum of Antiquities at Saint-Germain-en-Laye) between 1951 and 1955. Delporte concentrated his excavations in the areas immediately to the south of the main trench excavated through the front part of the site by Bailleau, in the area that now has the open form of a collapsed cavern. The detailed published reports^{16–18} and unpublished excavation records retained in the National Museum of Antiquities¹⁹ leave no doubt that Delporte’s excavations were carried out and recorded to high technical standards, in the manner of his later, better-known excavations at La Ferrassie, La Rochette and elsewhere. Delporte recorded that the site contained a sequence of at least eight archaeological levels, extending over a total depth of at least 2.5 m and containing at least five clearly defined levels of Chatelperronian occupation (levels B1–B5) (Fig. 1). These were underlain by at least 1.5 m of late Mousterian deposits (layers C1–C3), containing a typical Mousterian of Acheulian tradition industry, including at least three typical cordiform bifaces. Delporte reported that all the Chatelperronian levels were clearly stratified and marked by a succession of strongly reddened horizons that could be

¹Department of Archaeology, Cambridge University, Downing Street, Cambridge CB2 3DZ, UK. ²Oxford Radiocarbon Accelerator Unit, Oxford University, Keble Road, Oxford OX1 3QJ, UK.

traced as continuous, essentially 'sub-horizontal' units throughout the areas excavated. Secondary subdivisions could be recognized in three of these levels, and the top of layer B3 was marked by a thin line of flattened stones (Fig. 1). The clarity and regularity of the stratigraphy recorded in Delporte's excavations effectively excludes the possibility of any major stratigraphic disturbance during the formation of these levels. An intriguing feature of the Chatelperronian levels noted by Bailleau during his earlier excavations was a concentration of large mammoth tusks (up to two metres in length) apparently associated with a series of hearth deposits in the central part of the cave—a discovery that seems strongly reminiscent of the mammoth-tusk 'hut' structure recorded in the Chatelperronian levels of the Grotte du Renne at Arcy-sur-Cure, some 130 km to the north^{20,21}.

The archaeological material recovered from the Chatelperronian levels was analysed initially by Delporte himself⁸, and subsequently by F. Harrold²² and by one of the present authors (B.G.). The assemblage contains a total of over 750 artefacts, including over 200 retouched tools, distributed throughout levels B1–B5 and containing approximately 65 complete or fragmentary Châtelperron points, together with typical end scrapers, various forms of burins, truncated pieces and high frequencies (~85%) of typical blade technology. The richest levels were those of B5, B4 and B3 in the lower part of the sequence. Delporte noted that virtually all of the Chatelperronian material was manufactured from poor quality, pale brown flint apparently derived from alluvial sources close to the town of Tilly, 14 km to the northeast—a fact that probably explains the relatively small size of the artefacts, which included a number of unusually small specimens of Châtelperron points. Delporte

reported that diagnostically Mousterian forms were conspicuously absent from the assemblage.

Chatelperronian/Aurignacian interstratification

The most significant feature recorded during Delporte's excavation (as he himself emphasized) was the occurrence of a number of diagnostically Aurignacian tool forms stratified clearly within the Chatelperronian levels, and concentrated in level B4 in the lower part of the Chatelperronian sequence—that is, the level stratified between the rich Chatelperronian assemblages in levels B5 and B3 (Fig. 2). These pieces were individually described and illustrated in Delporte's report, and include at least three typical edge-retouched Aurignacian blades, two thick carinated scrapers and at least five small, inversely retouched 'Lamelle Dufour' bladelet forms. A large majority of these diagnostically Aurignacian artefacts were recovered from level B4, overlain by three further levels with typically Chatelperronian industries. Delporte further reported that at least four of the five Aurignacian blades and carinated scrapers were manufactured from either a unique, high quality local flint or from a distinctive, high quality blue-grey or pale brown 'Chalcedony' flint, which must have been imported into the site from distant flint sources at least 100 km to the north, and which contrasted sharply with the poor quality 'Tilly' flint from which virtually all of the Chatelperronian pieces were manufactured²³. The nineteenth-century excavations produced three further typical specimens of edge-retouched Aurignacian blades manufactured from the same distinctive, high-quality, imported Chalcedony flint (Fig. 2). Perhaps most significantly, Delporte also recovered two specimens of perforated animal teeth

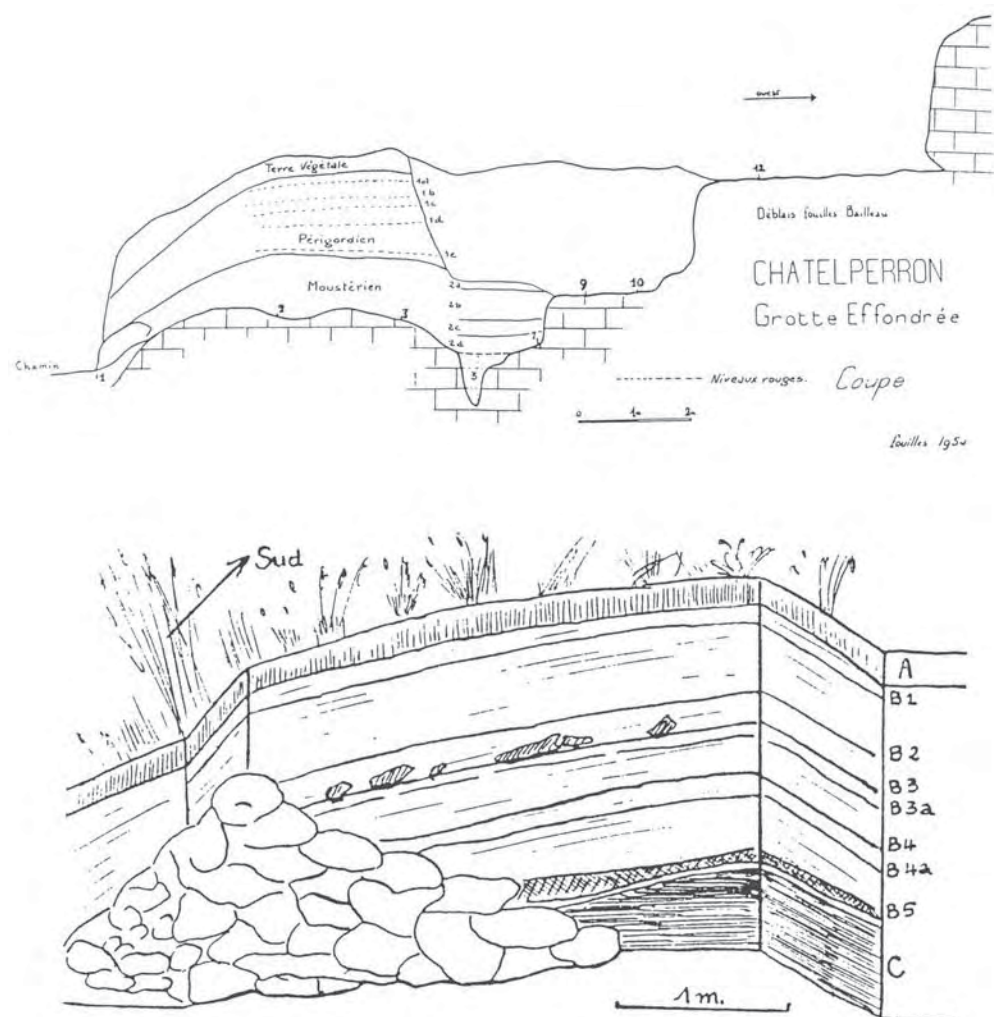


Figure 1 | Stratigraphy of the Grotte des Fées at Châtelperron. Sections recorded through the Chatelperronian levels (layers B1–B5) in Delporte's excavations at Châtelperron, immediately to the south of the nineteenth-century excavations in the front part of the cave^{17,18}. The upper diagram shows the main east–west section through the site, and the lower diagram shows the sections recorded in the smaller cutting excavated into the southern face of the long section. Note the regular, 'sub-horizontal' pattern of the stratigraphy and the thin layer of stones on top of layer B3. All of the Chatelperronian levels (layers B1–B5) were marked by thin horizons of strongly reddened sediments. The Chatelperronian levels were underlain by Mousterian levels (layer C) and overlain by the spoil of the nineteenth-century excavations (layer A). Reproduced from drawings by Delporte (refs 17, 18).

from level B4 (a fox canine and the canine of a large feline), which showed perforations produced in exactly the same way as those reported from many well-documented Aurignacian sites in France—that is, by initial scraping to thin the root of the tooth, followed by perforation of the hole²⁴ (Fig. 3). Finally, he reported that the nineteenth-century excavations had produced at least one typical specimen of a split-base bone or antler point (the diagnostic fossil type of the French early Aurignacian; specimen 10 in Fig. 2) together with at least three additional perforated teeth, but the exact stratigraphic provenance of these finds remains unknown.

The importance and potential implications of this interstratified sequence of Chatelperronian and Aurignacian horizons at the Grotte des Fées was clearly recognized by Delporte himself^{18,19} and has been commented on by several later workers^{7,24–27}, but it seems to have been overlooked in other recent discussions of these issues^{5,6,13,14}. The critical importance of this sequence rests on the fact that the Chatelperronian has been found at two other sites in France (Saint-Césaire in the Charente-Maritime and Arcy-sur-Cure in the Yonne Department, only 130 km to the north of Châtelperron) in apparently direct association with distinctively Neanderthal skeletal remains^{2,28,29}, whereas Aurignacian industries seem to be associated exclusively with anatomically modern populations in central and western Europe^{9,11,12}. The implication is clear that the site shows either a directly interstratified sequence of Neanderthal and anatomically modern human occupations, or at least a very close contact and interaction between these two populations within this particular region of France.

Radiocarbon dating

In view of the importance of the archaeological succession at Châtelperron, direct dating of the sequence seemed essential. The 13 bone samples submitted for radiocarbon dating were selected by one of the present authors (B.G.) from bags of faunal material retained at the National Museum of Antiquities at St Germain-en-Laye, labelled as deriving from level B5, B4 or levels B1–B3 of Delporte's excavations. The bone fragments were unwashed, had not been treated with preservatives, and consisted mainly of shaft fragments of fractured long bones of large herbivores. The bones were dated by accelerator mass spectrometry (AMS) at the Oxford Radiocarbon Accelerator Unit. The Oxford laboratory dates bone using an ultrafiltration technique to extract good quality collagen for dating (see Methods). This has been shown to eliminate contaminants of small molecular mass and often results in older, and we think more reliable, results compared with more routine methods³⁰.

The results of the AMS measurements are shown in Table 1. They reveal that, with one exception, the dates show a highly consistent pattern that agrees closely with the documented stratigraphic sequence of the samples. The three dates from the basal Chatelperronian level (B5) are effectively identical (between $40,650 \pm 600$ and $39,150 \pm 600$ yr BP), whereas the six samples from the combined later Chatelperronian levels (layers B1–B3) cluster between $36,340 \pm 320$ and $34,550 \pm 500$ yr BP (Fig. 4). The two samples from the intervening level B4 (which contained the Aurignacian material) produced dates of $39,780 \pm 390$ and $35,540 \pm 280$ yr BP.

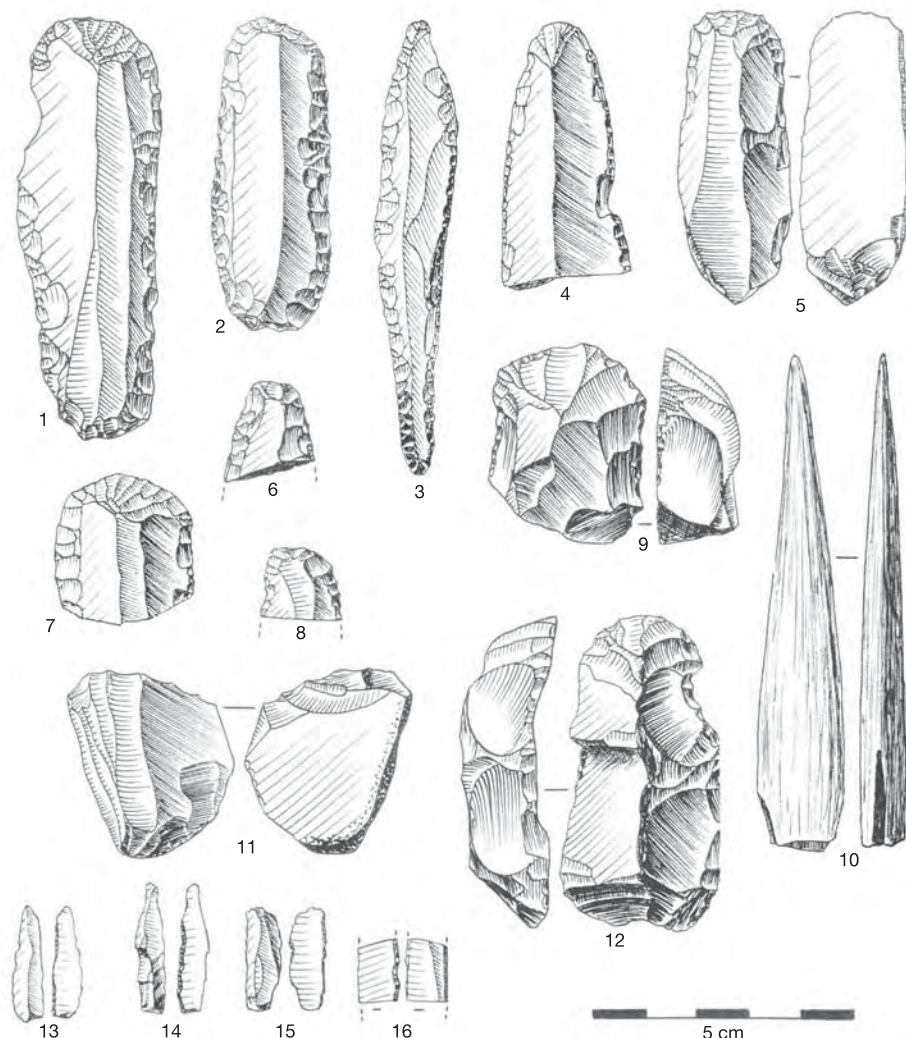


Figure 2 | Aurignacian artefacts from the Grotte des Fées at Châtelperron. Numbers 1–8, edge-retouched Aurignacian blades; 9 and 12, thick, carinate end scrapers; 11, bladelet core on large flake; 13–16, inversely retouched Lamelles Dufour bladelets; 10, split-base bone or antler point. Numbers 1, 6, 9, 11, 15 and 16 derive from level B4 of Delporte's excavations. Numbers 3, 4, 5, 7 and 10 derive from nineteenth-century excavations, for which no detailed stratigraphic provenance is available. Numbers 1, 2, 4–8, 11 and 13 are made on imported Chalcedony; 9 is made on high quality local flint; 3 is made on imported jasper. Adapted from drawings in ref. 19.



Figure 3 | Perforated animal teeth from level B4 of Delporte's excavation. Left, fox canine; right, canine of a large feline. In both specimens, the holes were produced by initial scraping to reduce the thickness of the root, followed by perforation of the hole. This is identical to the procedures used in the perforation of similar teeth from many Aurignacian sites in France^{24,27}. Several similar teeth were recovered from the nineteenth-century excavations at the site.

These two dates cluster with those from the immediately underlying and overlying Chatelperronian levels, respectively (Fig. 4), and apparently indicate the existence of a significant depositional hiatus during the formation of level B4. It was presumably during this interval that the brief episode of Aurignacian occupation on the site took place. The only aberrant date in the series is that of $>53,900$ yr BP, obtained for sample OxA-13620 from levels B1–B3, which is clearly inconsistent with the remainder of the dating for the site and presumably reflects an individual bone fragment that was either misplaced or mis-labelled during the original excavations (or possibly during post-excavation handling of the finds), or was stratigraphically derived in some way from the underlying Mousterian levels. Alternatively, this sample might have derived from the spoil of the nineteenth-century excavations, which directly overlay the final Chatelperronian level (B1) in Delporte's excavation, and was incorrectly attributed to the underlying level. Perhaps significantly, this is the only bone sample that shows signs of carnivore modification.

The dates as a whole clearly show that the basal Chatelperronian

level (B5) dates from approximately 39,000–40,000 yr BP in radiocarbon terms (see below), whereas the upper levels (B1–B3) apparently span the period $\sim 36,000$ –34,500 yr BP. These results are in good agreement with the dating of other Chatelperronian levels in France, which similarly span the range from approximately 40,000 to at least 34,000–35,000 yr BP (see Fig. 4). The implication is clear that the thin, intercalated level of Aurignacian material recorded in level B4 must date between approximately 36,000 and 39,000 yr BP in radiocarbon terms—slightly earlier than other dates for early Aurignacian levels in France, but comparable with those from adjacent areas of west-central Europe, such as Geissenklösterle and Keilberg Kirche in southern Germany and Willendorf II in Austria^{4,7,31,32}.

Although these dates are well beyond the range for which direct calibration is possible, they can be compared to the recently published radiocarbon record from the Cariaco Basin deep-sea core sequence³³. This record is tied to the timescale of the GISP2 Greenland ice core³⁴, for which we have useful climatic indicators, in particular the $\delta^{18}\text{O}$ ($^{18}\text{O}/^{16}\text{O}$ ratio) record^{35,36}. Recent research on combined oxygen isotope and pollen records in a number of deep-sea cores from adjacent to the Portuguese coast confirms that the climatic oscillations recorded in the GISP2 ice core record are in close agreement with those experienced in western Europe^{37,38}. In order to view the data here as a whole, we have treated the samples from B5, B4 and B1–B3 as three distinct groups (excluding OxA-13620). We have considered the sum of the probability distribution functions as an indication of the range of human activity in the three contexts. These summed distributions are shown in Fig. 5 (light grey). We have also applied a simple three-phase bayesian model using OxCal, with the three phases being B5, B4 and B1–B3 (which are undifferentiated). The distributions from this analysis are also shown in Fig. 5 (dark grey), including the dates for the transitions between phases.

What is clear from the Cariaco basin data is that the overall time range of the human occupation is likely to be much shorter than the range of the radiocarbon dates, with the occupation of B5 centred on 42,000–43,000 yr BP, B4 on $\sim 41,000$ –42,000 yr BP and that of B1–B3 on $\sim 40,000$ –41,000 yr BP (ages estimated from the GISP2 ice core)³⁴. Interestingly, the occupation in phases B5 and B1–B3 seems to be correlated with brief warmer spells visible in the $\delta^{18}\text{O}$ record (see Fig. 5). The two samples from B4 are somewhat different in age and correlate with the end of one warmer period and the start of the next, separated by an intervening colder phase. This coincides with the brief episode of Aurignacian occupation in level B4. A displacement of Aurignacian populations from central Europe to France in

Table 1 | Radiocarbon determinations of bone samples from the Grotte des Fées at Châtelperron

OxA	Sample ref.	^{14}C age (yr BP)*	C:N ratio†	$\delta^{13}\text{C}\ddagger$	$\delta^{15}\text{N}\S$	Sample size (mg)	Pretreatment yield (mg)	C (%)¶
Layers B1 to B3								
14165	B1/3A	36,340 $\pm$ 320	3.2	−18.9	7.9	1,040	30.7	44.5
13617	B1/3B	34,550 $\pm$ 500	3.3	−20.3	12.2	520	10.5	44.7
13618	B1/3C	35,890 $\pm$ 380	3.4	−18.5	7.2	600	19.0	43.9
13619	B1/3D	35,400 $\pm$ 450	3.4	−18.5	6.7	480	14.5	43.1
14166	B1/3E	34,940 $\pm$ 330	3.2	−18.8	5.9	1,000	21.3	42.9
13723	B1/3F	36,000 $\pm$ 1,000	3.3	−18.8	5.8	500	6.0	45.7
13620	B1/3G	$>53,900$	3.4	−20.4	5.8	580	19.7	45.6
13724	B1/3H	36,250 $\pm$ 750	3.3	−19.2	6.1	480	8.9	43.0
Layer B4								
14318	B4(1)	35,540 $\pm$ 280	3.2	−21.7	6.9	1,400	51.5	39.4
14319	B4(3)	39,780 $\pm$ 390	3.1	−20.7	7.7	1,100	62.9	41.5
Layer B5								
14320	B5(2)	39,240 $\pm$ 380	3.2	−21.5	6.7	1,240	58.0	39.6
13621	B5A	40,650 $\pm$ 600	3.3	−19.8	5.5	460	21.3	44.9
13622	B5B	39,150 $\pm$ 600	3.4	−21.0	7.6	440	16.2	44.1

*Radiocarbon ages are reported in yr BP (after ref. 45) with one standard deviation.

†C:N ratios ought to be within the range 2.9–3.5 for acceptance.

‡ $\delta^{13}\text{C} = [^{13}\text{C}/^{12}\text{C}]_{\text{sample}} / [^{13}\text{C}/^{12}\text{C}]_{\text{standard}} - 1$, and is reported in ‰ notation with respect to the Vienna Pee Dee Belemnite (VPDB) standard. Machine error for carbon is $\pm 0.2\text{‰}$.

§ $\delta^{15}\text{N} = [^{15}\text{N}/^{14}\text{N}]_{\text{sample}} / [^{15}\text{N}/^{14}\text{N}]_{\text{standard}} - 1$, and is reported in ‰ notation with respect to the AIR standard. Machine error for nitrogen is 0.3‰ for nitrogen.

||Pretreatment yield is in milligrams and refers to the weight of the lyophilized ultrafiltered gelatin extracted from the bone.

¶Per cent C is carbon upon combustion.

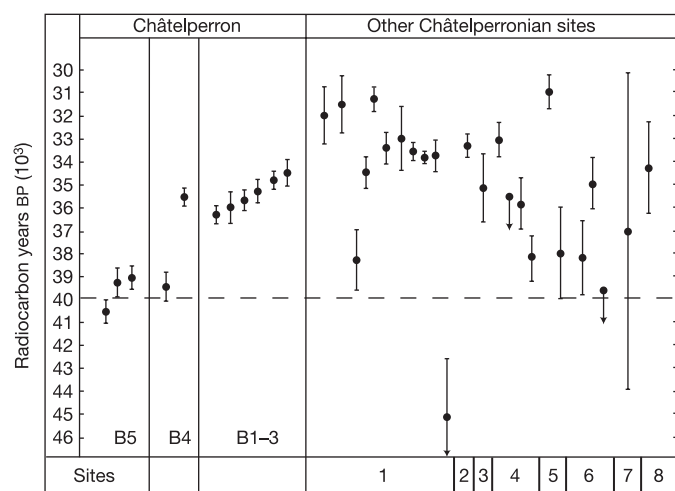


Figure 4 | Radiocarbon dates for Châtelperron and other Châtelperronian sites. Comparison of radiocarbon dates for levels B5, B4 and combined levels B1–B3 at Châtelperron with those from other Châtelperronian sites in France (sites 1–6) and northern Spain (sites 7, 8) (refs 7, 8, 46, 47). The dates are all shown in uncalibrated radiocarbon years before present, with error bars indicating one standard deviation. The sites are: 1, Arcy-sur-Cure; 2, Les Cottés; 3, Camiac; 4, Combe Saunière; 5, Roc de Combe; 6, Grotte XVI; 7, Cueva Morín; 8, Labeko Koba. Dates younger than ~33,000 yr BP are probably underestimates owing to sample contamination by more recent carbon.

response to sharply colder conditions at this time (especially perhaps the onset of severe winters) would hardly be surprising in ecological and demographic terms, as winter temperatures in the more oceanic areas of western Europe are likely to have been significantly milder than those further to the east³⁹. The same shift to colder conditions could well have caused a temporary displacement of Neanderthal territories towards the south, in effect leaving the region open to penetration by anatomically modern groups. The ensuing return to warmer conditions in the later Châtelperronian levels would have allowed the last Neanderthal groups to shift northwards again. If these climatic correlations are correct, the implication seems to be that the early anatomically modern Aurignacian populations in western Europe were rather better equipped to cope with cold climatic conditions than were the last Neanderthal groups.

Conclusions

In view of the doubts that have recently been expressed over the validity of other reported interstratifications of Châtelperronian and Aurignacian levels at two other sites in France (Roc de Combe and Le Piage in the Lot region, as well as at El Pendo in northern Spain)^{5,6,13,14}, it should be emphasized that the stratigraphic coherence of the radiocarbon dates obtained for the Grotte des Fées sequence reinforces the stratigraphic integrity of this succession. Clearly, if the material from the upper Châtelperronian levels (B1–B3) had been displaced or redeposited from the underlying Châtelperronian levels (B5) by some unexplained mechanisms, then the much younger dates secured on the six bone samples from levels B1–B3 would be impossible to explain (Fig. 4). The same conclusion seems to be equally inescapable from the very regular and clear-cut stratigraphic patterning of the upper Châtelperronian levels recorded in Delporte's excavations and published sections (Fig. 1). It is equally difficult to visualize how all of the distinctively Aurignacian artefacts (including two perforated teeth) could have travelled downwards through almost a metre of overlying stratified Châtelperronian levels and come to rest in layer B4. These data strongly support the chronological coexistence—and therefore potential demographic and cultural interactions—between the last Neanderthal and the earliest

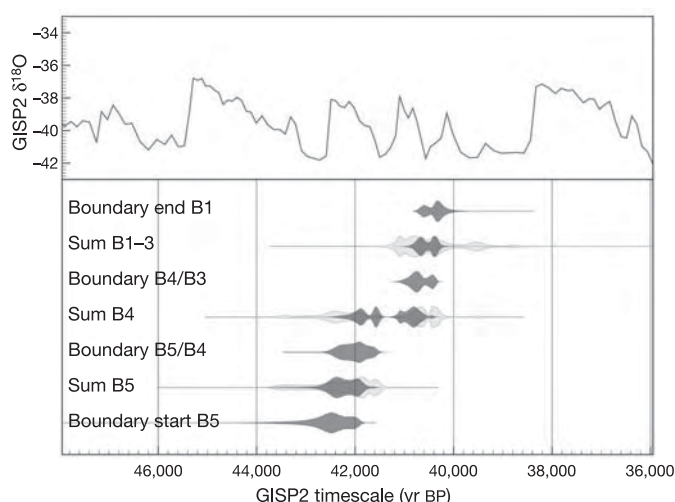


Figure 5 | Comparison of Châtelperron data to the radiocarbon record from the Cariaco basin³³. The light grey 'sum' distributions show the result of simple comparison (using the calibration procedures of OxCal⁴⁴); dark grey 'sum' distributions show the result of applying a three-phase bayesian model. The 'boundary' distributions show the start and end of each phase estimated from the model. The $\delta^{18}\text{O}$ record from the GISP2 Greenland ice core sequence is shown to give some indication of global climate trends over this time range.

anatomically and behaviourally modern human populations in western Europe.

METHODS

Radiocarbon dating. Samples of bone were prepared for AMS ^{14}C dating at the Oxford Radiocarbon Accelerator Unit. Bone was pre-cleaned using an aluminium oxide shotblaster. Routine collagen extraction procedures were applied⁴⁰ together with an additional ultrafiltration pretreatment⁴¹. This method allows only the >30 kDa molecular mass fraction of the soluble gelatin to be retained for radiocarbon assay. Lyophilized gelatin samples were combusted using a Roboprep CHN sample converter unit and analysed using a Europa Scientific 20-20 mass spectrometer operating in continuous flow mode. Graphite was prepared using routine methods⁴². We determined the preservation of the bone collagen using the carbon-to-nitrogen atomic ratio (C:N) and the percentage weight of collagen extracted from the bone. The addition of exogenous carbon atoms will increase the C:N ratio, and depending on the age and size of the contaminant, may result in errors in the AMS determinations. The Châtelperron bones were all within the acceptable C:N range of 2.9–3.5 (Table 1). Values for percentage weight collagen ranged from 1.2 to 4.6, which is poor, and in some instances (OxA-13723 and OxA-13724) only marginally above the minimum threshold for acceptance of 1% weight at Oxford. This indicates that ~6–23% of the collagen remains in the bone as reasonably intact protein.

The removal of low molecular mass (<30 kDa) material using ultrafiltration often yields older radiocarbon ages compared with other methods³⁰. This is because proteins of high molecular mass are retained, whereas degraded and broken up collagen, salts, and other low molecular mass contaminants (which might be of different ^{14}C age) pass through the ultrafilters.

The bayesian model used for data analysis was set up in OxCal^{43,44} with three phases. The dates are beyond the range for calibration, so the analysis was performed using the data from the Cariaco Basin for comparison³³ and the model was specified by OxCal commands detailed in the Supplementary Methods.

The dates of the boundaries and the sums of the dates for the three phases are plotted in Fig. 5. The sums give an indication of the temporal distribution of the dates within the phases, and the boundaries give estimates for the dates of transition between the phases.

Received 14 December 2004; accepted 6 July 2005.

Published online 31 August 2005.

1. Bar-Yosef, O. The Upper Paleolithic revolution. *Annu. Rev. Anthropol.* **31**, 363–393 (2002).
2. Hublin, J.-J. in *The Geography of Neandertals and Modern Humans in Europe and*

- the Greater Mediterranean (eds Bar-Yosef, O. & Pilbeam, P.) 157–182 (Peabody Museum, Harvard Univ., Cambridge, 2000).
3. Klein, R. G. Archaeology and the evolution of human behaviour. *Evol. Anthropol.* **9**, 7–36 (2000).
 4. Conard, N. & Bolus, M. Radiocarbon dating the appearance of modern humans and the timing of cultural innovations in Europe: new results and new challenges. *J. Hum. Evol.* **44**, 331–371 (2003).
 5. Zilhão, J. & d'Errico, F. The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neanderthal extinction. *J. World Prehist.* **13**, 1–68 (1999).
 6. Zilhão, J. & d'Errico, F. (eds) *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (Instituto Português de Arqueologia, Lisbon, 2003).
 7. Mellars, P. A. The Neanderthal problem continued. *Curr. Anthropol.* **40**, 341–364 (1999).
 8. Mellars, P. A. in *The Geography of Neandertals and Modern Humans in Europe and the Greater Mediterranean* (eds Bar-Yosef, O. & Pilbeam, P.) 35–48 (Peabody Museum, Harvard Univ., Cambridge, 2000).
 9. Mellars, P. A. Neanderthals and the modern human colonization of Europe. *Nature* **432**, 461–465 (2004).
 10. Mellars, P. A. The impossible coincidence: a single species model for the origins of modern human behaviour in Europe. *Evol. Anthropol.* **14**, 12–27 (2005).
 11. Gambier, D. in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans* (eds Mellars, P. A. & Stringer, C.) 194–211 (Edinburgh Univ. Press, Edinburgh, 1989).
 12. Churchill, S. & Smith, F. Makers of the early Aurignacian of Europe. *Yb. Phys. Anthropol.* **43**, 61–115 (2000).
 13. d'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pelegrin, J. Neanderthal acculturation in western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* **39**, S1–S44 (1998).
 14. Bordes, J. G. in *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (eds Zilhão, J. & d'Errico, F.) 223–246 (Instituto Português de Arqueologia, Lisbon, 2003).
 15. Breuil, H. Etudes de morphologie paléolithique. 2. L'industrie de la grotte de Châtelperron (Allier) et d'autres gisements similaires. *Revue Ecole Anthropol. Paris* **29**, 29–40 (1910).
 16. Delporte, M. H. L'industrie de Châtelperron et son extension géographique. *Congrès Préhist. Fr.* **14**, 233–250 (1955).
 17. Delporte, M. H. Les fouilles des grottes paléolithiques de Châtelperron (Allier). *Gallia* **13**, 79–84 (1955).
 18. Delporte, M. H. La Grotte des Fées de Châtelperron (Allier). *Congrès Préhist. Fr.* **15**, 452–477 (1957).
 19. Delporte, M. H. *Excavations at Châtelperron* (Musée des Antiquités Nationales, Paris, unpublished manuscript ~1964).
 20. Leroi-Gourhan, A. Les fouilles d'Arcy-sur-Cure (Yonne). *Gallia Préhistoire* **4**, 3–16 (1961).
 21. Farizy, C. in *Paléolithique Moyen Récent et Paléolithique Supérieur Ancien en Europe* (ed. Farizy, C.) 281–289 (Mémoires du Musée de Préhistoire d'Ile de France 3, Paris, 1990).
 22. Harrold, F. B. A study of the Châtelperronian PhD Thesis, Univ. Chicago (1978).
 23. Delporte, M. H., Surlim, F. & Urgal, A. *Châtelperron: Un Grand Gisement Préhistorique de l'Allier* (Conseil Général de l'Allier, Moulins, 1999).
 24. White, R. Personal ornaments from Grotte du Renne at Arcy-sur-Cure. *Athena Rev.* **2**, 41–46 (2001).
 25. Combier, J. in *Paléolithique Moyen Récent et Paléolithique Supérieur Ancien en Europe* (ed. Farizy, C.) 267–278 (Mémoires du Musée de Préhistoire d'Ile de France 3, Paris, 1990).
 26. Pelegrin, J. *Technologie Lithique: Le Châtelperronien de Roc-de-Combe (Lot) et de la Côte (Dordogne)* (CNRS, Paris, 1995).
 27. d'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pelegrin, J. Neanderthal acculturation in western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* **39**, S1–S44 (1998).
 28. Lévêque, F. & Vandermeersch, B. Découverte de restes humains dans un niveau castelperronien à Saint-Césaire (Charente-Maritime). *C.R. Acad. Sci. Paris* **129**, 187–189 (1980).
 29. Hublin, J.-J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature* **381**, 224–226 (1996).
 30. Bronk Ramsey, C., Higham, T., Bowles, T. & Hedges, R. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* **46**, 155–163 (2004).
 31. Richter, D., Waiblinger, J., Rink, J. & Wagner, G. Thermoluminescence, electron-spin-resonance and ^{14}C dating of the late Middle and early Upper Paleolithic site of Geissenklösterle Cave in southern Germany. *J. Archaeol. Sci.* **27**, 71–89 (2000).
 32. Haesaerts, P. & Teyssandier, N. in *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (eds Zilhão, J. & d'Errico, F.) 133–151 (Instituto Português de Arqueologia, Lisbon, 2003).
 33. Hughen, K. et al. ^{14}C activity and global carbon cycle changes over the past 50,000 years. *Science* **303**, 202–207 (2004).
 34. Meese, D. A. et al. The Greenland Ice Sheet Project 2 depth-age scale: methods and results. *J. Geophys. Res.* **102** (C12), 26411–26423 (1997).
 35. Grootes, P., Stuiver, M., White, J., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
 36. Grootes, P. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10^3 to 10^5 -year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).
 37. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
 38. Voelker, A. H. L. & workshop participants, Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: a database. *Quat. Sci. Rev.* **21**, 1185–1212 (2002).
 39. Van Andel, T. H. & Davies, W. (eds) *Neanderthals and Modern Humans in the European Landscape during the Last Glaciation* (McDonald Institute for Archaeological Research, Cambridge, 2003).
 40. Bronk Ramsey, C., Pettitt, P., Hedges, R., Hodgins, G. & Owens, D. Radiocarbon dates from the Oxford AMS system: Archaeometry Datelist 30. *Archaeometry* **42**, 459–479 (2000).
 41. Bronk Ramsey, C., Higham, T. & Leach, P. Towards high precision AMS: progress and limitations. *Radiocarbon* **46**, 17–24 (2004).
 42. Bronk Ramsey, C. & Hedges, R. Hybrid ion sources: Radiocarbon measurements from microgram to milligram. *Nucl. Instrum. Methods Phys. Res.* **123**, 539–545 (1999).
 43. Bronk Ramsey, C. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* **37**, 425–430 (1995).
 44. Bronk Ramsey, C. Development of the radiocarbon program OxCal. *Radiocarbon* **43**, 355–363 (2001).
 45. Stuiver, M. & Polach, H. Discussion: Reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
 46. David, F. et al. Le Châtelperronien de la Grotte du Renne à Arcy-sur-Cure (Yonne). Données sédimentologiques et chronostratigraphiques. *Bull. Soc. Préhist. Fr.* **98**, 207–230 (2001).
 47. Lucas, G., Rigaud, J.-P., Simek, J. & Soressi, M. in *The Chronology of the Aurignacian and of the Transitional Technocomplexes: Dating, Stratigraphies, Cultural Implications* (eds Zilhão, J. & d'Errico, F.) 289–300 (Instituto Português de Arqueologia, Lisbon, 2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to the Musée des Antiquités Nationales, Saint-Germain-en-Laye, for providing the samples for the radiocarbon dating, to the NERC for their support for the Oxford Radiocarbon Accelerator Dating Service (ORADS) at Oxford, to D. Kemp for assistance with the illustrations, to J.-P. and D. Steenhuyse, M. C. Dawson, C. Schwab, R. White and N. Ashton for other assistance, and to K. Boyle, R. Hedges, T. Higham and N. Shackleton for discussions. B.G. is indebted to Magdalene College, Cambridge, for a Leslie Wilson Research Studentship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.M. (pam59@cam.ac.uk).

The yeast Pif1p helicase removes telomerase from telomeric DNA

Jean-Baptiste Boulé^{1*}, Leticia R. Vega^{1*†} & Virginia A. Zakian¹

Telomeres are the physical ends of eukaryotic chromosomes. Genetic studies have established that the baker's yeast Pif1p DNA helicase is a negative regulator of telomerase, the specialized reverse transcriptase that maintains telomeric DNA, but the biochemical basis for this inhibition was unknown. Here we show that *in vitro*, Pif1p reduces the processivity of telomerase and releases telomerase from telomeric oligonucleotides. The released telomerase is enzymatically active because it is able to lengthen a challenger oligonucleotide. *In vivo*, overexpression of Pif1p reduces telomerase association with telomeres, whereas depleting cells of Pif1p increases the levels of telomere-bound Est1p, a telomerase subunit that is present on the telomere when telomerase is active. We propose that Pif1p helicase activity limits telomerase action both *in vivo* and *in vitro* by displacing active telomerase from DNA ends.

Saccharomyces cerevisiae telomeres consist of ~350 base pairs of the sequence C₁₋₃A/TG₁₋₃. As in most eukaryotes, yeast telomeric DNA is elongated by telomerase, a specialized reverse transcriptase. Yeast telomerase contains the catalytic subunit (Est2p), the templating RNA (TLC1) and additional protein subunits, such as Est1p, that participate in the recruitment or activation of the enzyme.

From yeasts to humans, telomerase activity is highly regulated. Yeast telomerase can lengthen telomeres only in late S/G2 phase^{1,2}, even though the catalytic core of telomerase is telomere-bound throughout most of the cell cycle^{3,4}. In contrast, Est1p is telomere-associated only in late S/G2 phase, the time of telomerase action³⁻⁵.

Pif1p is a 5' to 3' helicase that inhibits telomerase-mediated telomere lengthening^{6,7} and *de novo* telomere addition^{6,8,9}. In the absence of Pif1p, telomeres increase in length by ~100 base pairs, whereas Pif1p overexpression results in an ~80-base-pair decrease in telomere length^{6,7}. Lack of Pif1p results in a ~600-fold increase in *de novo* telomere addition at spontaneous double-strand breaks⁶, and a ~1,000-fold increase in the generation of the types of gross chromosomal rearrangements that occur during tumour formation by addition of telomeric DNA to spontaneous chromosome breaks⁹. Cells expressing the catalytically-inactive Pif1p^{K264A} have the same phenotype as *pif1Δ* cells, and cells overexpressing Pif1p^{K264A} do not exhibit telomere shortening. Thus, the catalytic activity of Pif1p is required for its effects on telomeres. As Pif1p is telomere-associated *in vivo*, it probably affects telomerase directly⁷.

There is increasing evidence that positive and negative regulation of telomerase contributes to genome integrity. Here we employ *in vitro* and *in vivo* assays to show that Pif1p acts catalytically to remove telomerase from telomeric DNA.

Specificity of Pif1p helicase activity

We purified the nuclear forms of both wild-type Pif1p and the helicase-inactive Pif1p^{K264A} (ref. 7) (Fig. 1a) and tested their ability to unwind 20-base oligonucleotide DNA/DNA, RNA/RNA or RNA/DNA duplexes of identical nucleotide sequence, except that U replaced T in the RNA strands (Fig. 1b; see Methods for sequence of the helicase substrate). All substrates contained the same 20

nucleotide single-stranded 5' overhang, which is required for Pif1p unwinding¹⁰. As expected, recombinant Pif1p unwound the DNA/DNA substrate (Fig. 1b, lane 2). Pif1p also unwound the RNA/DNA hybrid in which the strand containing the single-stranded 5' overhang was made of DNA (Fig. 1b, lane 4) but not when it was made of RNA (Fig. 1b, lane 6), nor did it unwind an RNA duplex (Fig. 1b, lane 8). As assessed by bandshift experiments, Pif1p bound to single-stranded DNA but not significantly to single-stranded RNA (data not shown). Therefore, Pif1p requires a deoxyribose backbone to load onto its substrate. As expected, Pif1p^{K264A} had no unwinding activity (Fig. 1b, lanes 9–12), although it bound efficiently to single-stranded DNA (see Supplementary Fig. S2).

Pif1p reduces telomerase processivity

Processivity describes an enzyme's ability to undergo multiple reaction cycles without being released from its substrate. Two types of processivity have been defined for telomerase, nucleotide addition processivity and repeat addition processivity¹¹. The first refers to the ability of telomerase to copy the entire template region of the RNA

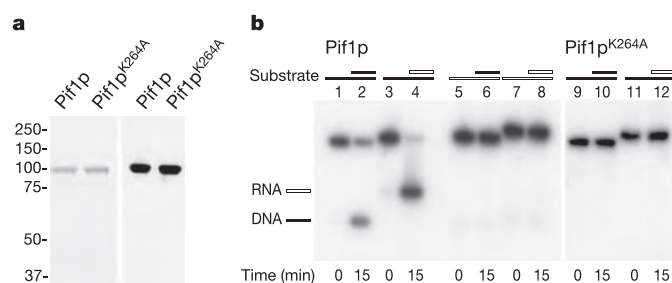


Figure 1 | Activity of Pif1p. **a**, Purified wild-type Pif1p and Pif1p^{K264A} were analysed by Coomassie staining (left) or western blotting (right). **b**, Radiolabelled substrates were incubated with 60 nM Pif1p (left) or Pif1p^{K264A} (right). At 0 and 15 min after ATP addition, aliquots were withdrawn and examined by electrophoresis. RNA strands are represented with open rectangles and DNA strands are represented by filled rectangles.

¹Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [†]Present address: Barry University, School of Natural and Health Sciences, Miami Shores, Florida 33161, USA.

*These authors contributed equally to this work.

without dissociating from the DNA primer. The second refers to the ability of the enzyme to translocate the RNA to the end of the DNA product after completing a round of template replication, allowing the enzyme to engage in a second round of RNA template replication without dissociating from the DNA (refs 11–13). Although human and ciliate telomerases display strong repeat addition processivity *in vitro*^{14,15}, in many organisms, including yeast, telomerase is fairly non-processive at both the nucleotide and repeat addition levels^{16–20}. Yeast telomerase stalls at each nucleotide position, resulting in a ladder of nucleotide addition products (for example, Fig. 2a, lane 1). Typically, no more than one round of RNA template replication is observed¹⁶. This lack of processivity is not due to dissociation of telomerase from the DNA primer as the enzyme remains tightly associated with the elongated product²¹.

If Pif1p dissociates telomerase RNA from telomeric DNA, it might limit nucleotide addition processivity of telomerase-mediated extension *in vitro*. To test this possibility, we added recombinant wild-type Pif1p or Pif1p^{K264A} to a standard primer extension telomerase assay. We partially purified telomerase from a strain overexpressing Est2p and TLC1 by fractionation of a yeast extract on a DEAE sepharose column^{16,21,22}. We used the standard telomeric primer Tel15 (5'-TGTGGTGTGTGGG-3') as substrate¹². Because Tel15 ends with three G bases, it anneals to the TLC1 telomerase RNA at a unique position (nucleotides A484 to C476). This unique annealing behaviour allows precise prediction of the sequence of the telomerase reaction product (TGTGGTG from +1 to +7) and quantification of the product made at each nucleotide position.

Total incorporation was fivefold higher in the presence of Pif1p, whereas addition of Pif1p^{K264A} had no effect on the amount of synthesis (Fig. 2a, b). Pif1p also affected the length distribution of reaction products (Fig. 2a, c, d). In the absence of Pif1p, the most common elongation products were +5 and +7; whereas in the presence of wild-type Pif1p, the most common elongation product was +2. As the amount of incorporation and distribution of reaction products were not affected by adding Pif1p^{K264A}, these effects require the enzymatic activity of Pif1p. In addition, because Pif1p^{K264A} bound to DNA as well as wild-type Pif1p (see Supplementary Fig. S2), the effects of wild-type Pif1p on the reaction cannot be explained by a competition between Est2p and Pif1p for binding to Tel15.

We determined the relative amounts of the elongation products at each extension position (Fig. 2c), and from these data processivity was expressed as the fraction of elongation products of a given length that continued to be elongated¹² (Fig. 2d). In the absence of Pif1p or in the presence of Pif1p^{K264A}, Tel15 primers that had been extended by 1, 2, 3 or 4 bases had a 90–100% probability of adding an additional base (Fig. 2d). However, in the presence of Pif1p, only 50% of the Tel15 primers that had been extended by 1 base were further lengthened; only 25% of the Tel15 primers that were lengthened by two bases were lengthened again. Thus, Pif1p had two important effects on telomerase activity *in vitro*. First, it increased the overall telomerase activity in the extract (Fig. 2b). Second, it reduced the length of the addition products (Fig. 2c). These effects were also observed at lower concentrations of Pif1p. The effect of

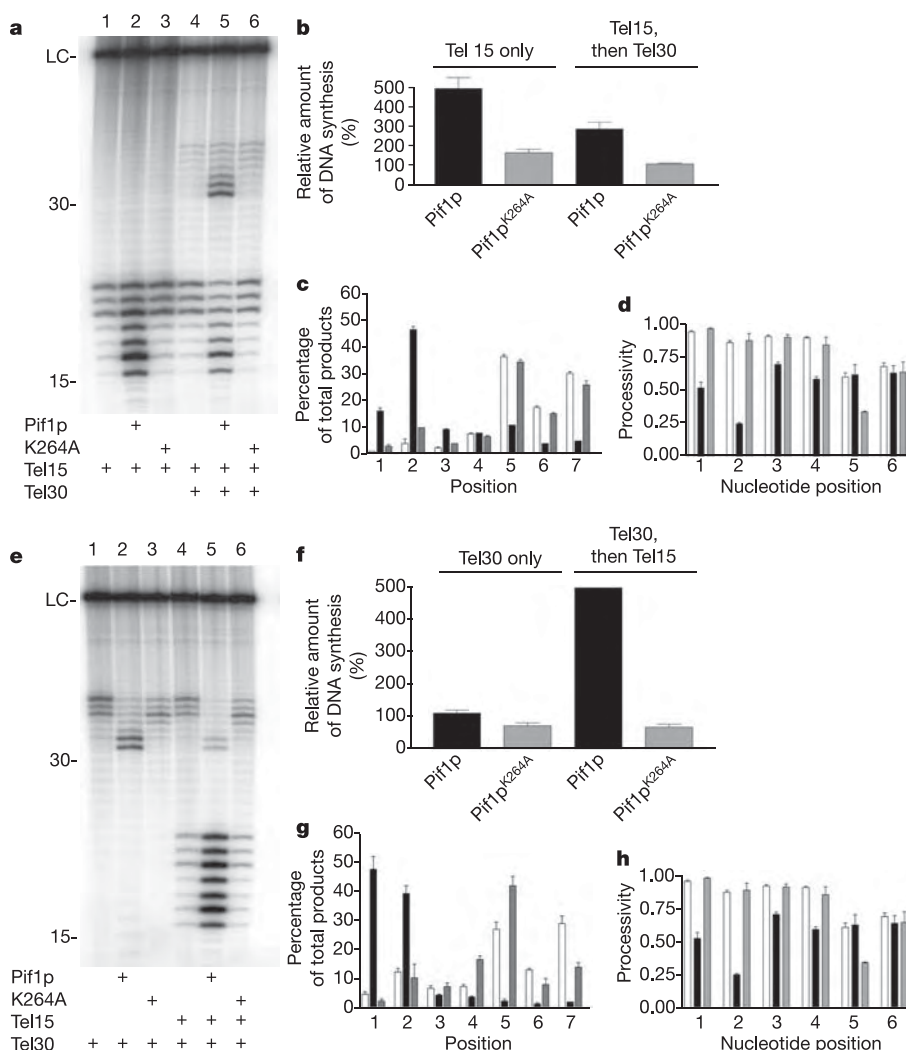


Figure 2 | Pif1p reduces telomerase processivity and promotes lengthening of challenger primers.

a, Partially purified telomerase was incubated with 0.1 μ M Tel15 and 1 μ M Pif1p or Pif1p^{K264A} (labelled K264A). When indicated, 1 μ M Tel30 was added 15 min after the beginning of the reaction. Reaction products were analysed by electrophoresis (LC, loading control). **b**, Quantification of the total DNA synthesis relative to synthesis in absence of Pif1p (Error bars, standard deviations from three independent experiments). **c**, **d**, Relative product distribution (**c**) and processivity (**d**) of telomerase in reactions carried out without addition of Tel30 with telomerase only (white bars), in presence of 1 μ M Pif1p (black bars) or 1 μ M Pif1p^{K264A} (grey bars). **e**–**h**, Similar to **a**–**d** except that reactions were begun with Tel30 and Tel15 was added after 15 min where indicated.

Pif1p addition is particularly noticeable at the +2 addition product, which becomes the main product at an equimolar [Pif1]:[TEL15] ratio (see Supplementary Fig. S3). Pif1p also affected nucleotide addition processivity when a 30-base oligonucleotide primer Tel30 (5'-CGCCATGCTGATCCGTGTGGTGTGTGGG-3') was used. The most common reaction product with Tel30 in the absence of Pif1p or in the presence of Pif1p^{K264A} was +5 or +7 bases; whereas in the presence of Pif1p, the most common elongation products were +1 and +2 (Fig. 2g). As with Tel15, in the absence of Pif1p or in the presence of Pif1p^{K264A}, primers that had been extended by 1 to 4 bases had a probability of close to 100% of further lengthening (Fig. 2h). In the presence of Pif1p, only 50% of the +1 Tel30 primers and 25% of the +2 Tel30 primers were further lengthened. Therefore, Pif1p reduces the nucleotide addition processivity of telomerase *in vitro*.

Removal of telomerase from DNA primers

In addition to its effects on nucleotide addition processivity, Pif1p increased the number of telomerase extension reactions that occurred. For example, using Tel15, the addition of wild-type Pif1p resulted in five times more total incorporation (Fig. 2b, Tel15 only). With Tel30, the total incorporation was comparable in the presence or absence of Pif1p (Fig. 2f, Tel30 only). However, as Pif1p reduced the average length of extension of Tel30 from 5–7 bases to 1–2 bases (Fig. 2g), there must have been more cycles of elongation in the presence of Pif1p.

How can an inhibitor of lengthening increase the total number of elongation reactions? In the typical primer extension reaction, yeast telomerase remains stably bound to the elongated primer²¹. The effects of Pif1p could be explained if Pif1p displaces telomerase in an active form, freeing it to lengthen a challenger primer. To test this possibility, telomerase was allowed to extend Tel15 for 15 min and then a tenfold excess of Tel30 was added. After 1 h, the reaction products were examined (Fig. 2a, lanes 4–6). In the absence of Pif1p or in the presence of Pif1p^{K264A}, there was very little lengthening of Tel30 (Fig. 2a, lanes 4 and 6). However, when Pif1p was added, lengthening of the Tel30 challenger primer was readily detected (Fig. 2a, lane 5). The amount of incorporation for Tel30 was three times greater in the presence of Pif1p than in its absence (Fig. 2b, Tel15, then Tel30). A similar result was seen when Tel30 was added first and Tel15 was used as the challenger primer (Fig. 2e, lanes 4–6). In the presence of Pif1p, there was five times more elongation of the challenger Tel15 primer (Fig. 2f, Tel30, then Tel15).

To provide additional evidence that Pif1p is acting by removing telomerase, we asked if Est2p is released from the Tel30 primer on Pif1p treatment. Tel30 was bound to magnetic beads and added to the standard telomerase reaction. After washing to remove unbound Est2p, Pif1p was added. Again the beads were washed and the supernatant tested for the presence of Est2p. Est2p was readily detected in the supernatant. This release required ATP and the enzymatic activity of Pif1p, consistent with the explanation that the effects of Pif1p on telomerase activity are due to the removal of Est2p from DNA ends (see Supplementary Fig. S4).

Removal of telomerase from chromosome ends

If Pif1p affects telomeres by removing telomerase from chromosome ends, Pif1p overexpression, which results in telomere shortening⁷, might reduce the level of Est2p and Est1p at telomeres. We overexpressed the nuclear form of either Pif1p or Pif1p^{K264A} by placing their expression under the control of the strong galactose-inducible *GAL1* promoter in strains expressing Myc18-Est2p, Myc9-Est1p or Myc9-Cdc13p (refs 3, 4). Each of the three epitope-tagged strains was transformed with an empty plasmid (control) or the same plasmid containing Gal-*PIF1* or Gal-*PIF1*^{K264A}. For each experiment, at least three independent transformants were analysed by chromatin immunoprecipitation (ChIP)⁴. We monitored association of the tagged protein with the right telomere of chromosome VI (VI-R)

and the left telomere of chromosome VII (VII-L), as well as the ability to immunoprecipitate the epitope-tagged proteins. Using these methods, in the control strain, telomeric DNA was enriched about 20-fold in the anti-Est2p immunoprecipitate (18-fold, VI-R; 24-fold, VII-L). When Pif1p was overexpressed, the amount of telomeric DNA in the anti-Est2p immunoprecipitate decreased almost twofold (1.8-fold, VI-R; 1.5-fold, VII-L). These differences are significant by the criterion of a *t*-test ($P = 0.0028$ for VI-R and 0.0001 for VII-L). In contrast, compared to vector alone, there was no significant difference in Est2p association with telomeres on overexpression of Pif1p^{K264A} ($P = 0.3726$ for VI-R; 0.05293 for VII-L) (Fig. 3a, b).

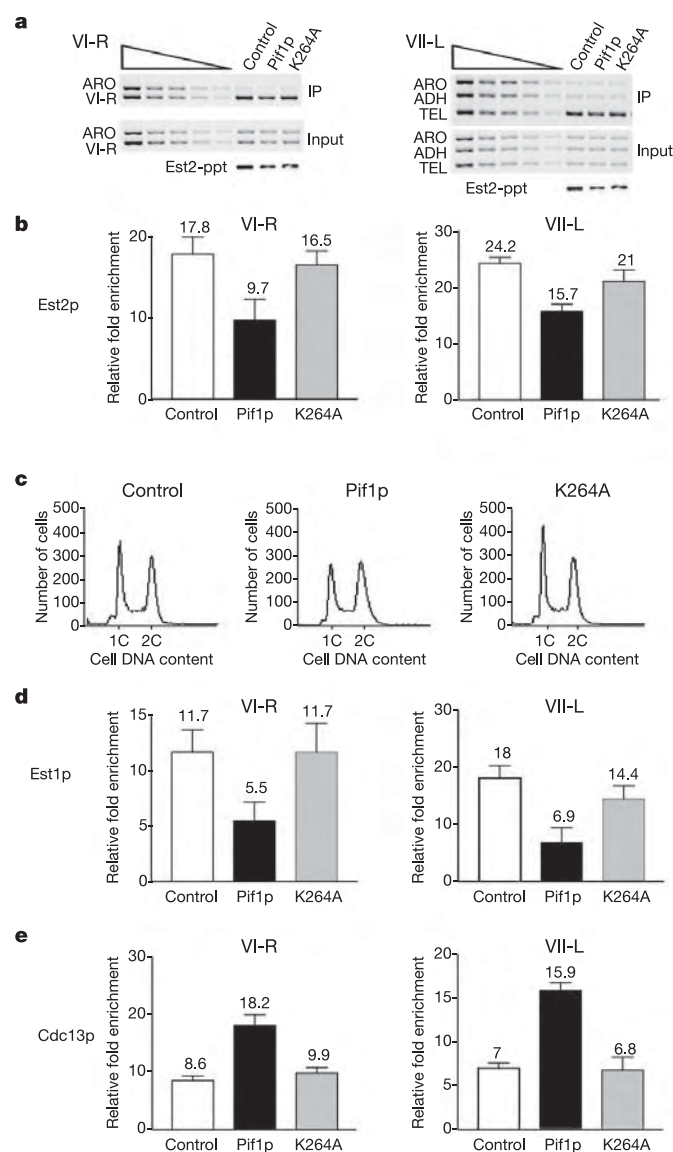


Figure 3 | Pif1p overexpression decreases Est2p and Est1p telomere association. **a–c**, Galactose grown cells expressing Myc18-Est2p with empty vector (control), Gal-wt *PIF1* (Pif1p) or Gal-*PIF1*^{K264A} (K264A) were analysed by ChIP (**a**) and FACS (**c**). Quantitative multiplex PCR was carried out using primer pairs specific for telomeric (VI-R, TEL), subtelomeric (ARO) or non telomeric (ARO) chromosomal loci. Amounts of VI-R (left) and VII-L (right) telomeric DNA in the anti-Myc IP and whole cell extract (input). Anti-Myc western of the immunoprecipitate (Est2-ppt). Quantitative analysis of telomeric association of Myc18-Est2p (**b**). Data expressed as average relative-fold enrichment $\pm$ standard deviation of telomeric band over ARO signal normalized to input DNA (ref. 4). FACS analysis (**c**) of cells for the ChIP analysis in **a**, **d**, **e**, Association of Myc9-Est1p (**d**) and association of Myc9-Cdc13p (**e**) with VI-R and VII-L telomeres as described for Myc18-Est2p.

Overexpression of wild-type Pif1p also caused significantly lower Est1p association with both the VI-R (2.1-fold) and VII-L (2.6-fold) telomeres ($P = 0.0134$ for VI-R; 0.0018 for VII-L; Fig. 3d). In contrast, overexpression of Pif1p^{K264A} did not have a significant effect on Est1p association ($P = 0.994$, VI-R; $P = 0.103$, VII-L). The reduced association of Est2p and Est1p with telomeres is specific as the level of telomere binding of Cdc13p was not reduced on Pif1p overexpression (Fig. 3e). Rather when wild-type Pif1p was overexpressed, there was significantly more telomeric DNA in the anti-Cdc13p immunoprecipitate (2.1-fold at VI-R, $P = 0.001$; 2.3-fold at VII-L, $P = 0.0001$). This increase in Cdc13p binding probably reflects reduced telomere protection and hence increased C-strand degradation as a result of removal of Est2p. When Pif1p^{K264A} was overexpressed, Cdc13p binding was indistinguishable from the vector control (VI-R, $P = 0.161$; VII-L, $P = 0.890$).

We also examined the effects of reduced Pif1p expression on telomere binding of Est2p, Est1p and Cdc13p in *PIF1* and *pif1Δ* cells. Although there was no significant change in the telomere association of Est2p or Cdc13p in *pif1Δ* cells, there were significantly higher levels of Est1p at both the VI-R (2.1-fold) and VII-L (1.6-fold) telomeres in the *pif1Δ* isolates ($P = 0.00279$ VI-R; 0.000114 VII-L; see Supplementary Fig. S5).

Discussion

We show that the *S. cerevisiae* Pif1p helicase, previously found to limit telomerase action *in vivo*^{6–9}, reduces the nucleotide addition processivity of telomerase *in vitro* (Fig. 2d, h). By both *in vitro* (Fig. 2; see also Supplementary Fig. S3) and *in vivo* (Figs 3 and 4) assays, Pif1p removed telomerase from DNA ends. *In vitro*, where the concentration of primers was high, the released telomerase resulted in an increased ability to lengthen a challenger primer (Fig. 2a, e, lane 5) and increased telomerase activity (Fig. 2b, f). Because *in vivo* the concentration of telomeres is much lower, Pif1p-mediated release of telomerase limits, rather than promotes, telomerase-mediated telomere lengthening^{6,7}. Both *in vitro* and *in vivo*, telomerase removal required the helicase activity of Pif1p (Fig. 2; see also Supplementary

Fig. S4). Because catalytically-inactive Pif1p^{K264A} bound even better to single-stranded DNA than wild-type Pif1p (see Supplementary Fig. S2), the effects of Pif1p on telomerase can not be explained by a model in which Pif1p and telomerase compete for binding to the DNA substrate. As Pif1p did not unwind RNA duplexes (Fig. 1), Pif1p is unlikely to affect telomerase processivity by remodelling telomerase RNA. However, Pif1p did unwind RNA/DNA hybrids *in vitro* (Fig. 1). Thus, Pif1p might displace telomerase by unwinding the short telomerase RNA/telomeric DNA hybrid that is the intermediate in the telomerase reaction. Alternatively (or in addition), Pif1p might directly remove Est2p from telomeric DNA^{23–26} or it might disrupt the interaction of DNA with the telomerase anchor site, a DNA binding site distinct from the catalytic site that has been proposed to hold telomerase at the telomere during translocation^{27,28}.

The ability of Pif1p to remove telomerase from DNA is unlikely to be a general property of yeast helicases. For example, mutations in the 134 other potential yeast helicases²⁹ were not found in the screen that identified multiple *pif1* alleles by their positive effects on *de novo* telomere addition (ref. 6 and unpublished data). Likewise, there were no other helicases whose mutation increased the recovery of gross chromosomal rearrangements by increasing telomere addition⁹.

In vivo, overexpression of Pif1p resulted in reduced Est2p and Est1p at telomeres (Fig. 3). This reduction occurred at two telomeres and was specific, as Cdc13p binding did not decrease but instead increased when Pif1p was overexpressed (Fig. 3). Likewise, deletion of Pif1p resulted in a significant ~twofold increase in Est1p telomere binding at the VI-R and VII-L telomeres (Fig. 4). However, there was no change in Est2p or Cdc13p telomere binding in *pif1Δ* cells (see Supplementary Fig. S5), even though Est2p telomere binding is *TLC1* dependent^{3,4}. Unlike Est1p, which binds telomeres only in late S/G2 phase, Est2p and Cdc13p are telomere-associated throughout most of the cell cycle^{3,4}. *In vivo*, only ~7% of yeast telomeres are lengthened by telomerase in a given cell cycle³⁰. We speculate that most telomeres are Est2p associated and that Est1p binds only the subset of telomeres that are lengthened by telomerase. If most telomeres are already Est2p associated, this fraction cannot be increased on Pif1p loss.

Pif1-like proteins are present in all multicellular eukaryotes whose genomes have been completely sequenced^{31,32}. Although there is no evidence as yet that these other Pif1-like proteins affect telomerase, natural catalytic inhibitors of telomerase might provide targets or models for anti-telomerase strategies.

METHODS

Helicase assay. For helicase assays, 1 μM top strand oligonucleotide (5'-CGCCATGGTGATCCGAGTGC-3' for the DNA substrate) was ³²P-labelled with T4 polynucleotide kinase, mixed with equimolar amounts unlabelled bottom strand oligonucleotide (5'-CACTGGCGTCTTACGGTCGGCACTCG-GATCACCATTGGCG-3' for the DNA substrate), heated to 95 °C, cooled overnight to 4 °C and gel purified. 1 nM gel purified substrate was incubated in the presence of the indicated amount of purified Pif1p or Pif1p^{K264A} (see Supplementary Methods for details of Pif1p purification) in buffer containing 20 mM Tris-HCl at pH 7.5, 50 mM NaCl (the optimal concentration for unwinding of both DNA/DNA and RNA/DNA hybrids), 5 mM MgCl₂, 100 μg ml⁻¹ BSA and 2 mM DTT with an ATP regeneration system (50 mM creatine phosphate and 50 μg ml⁻¹ creatine phosphokinase). Products were analysed on 12% non-denaturing gels and quantified on a PhosphorImager (Molecular Dynamics).

Telomerase extracts preparation and activity assay. For preparation of telomerase extracts, full-length *EST2* and *TLC1* were cloned into pESC-Trp (Stratagene) under the control of *GAL1* and *GAL10* promoters, respectively. *EST2* was in fusion with an amino-terminal HAT tag that did not affect telomere length (data not shown). The plasmid was introduced into a type II survivor from an *est1* protease-deficient strain (derived from BCY123, ref. 33), and overexpression achieved as described³⁴. Telomerase extracts were prepared by DEAE fractionation²² (see also Supplementary Methods). The extract from *Est2p/TLC1* overexpressing cells had five- to tenfold more activity per microlitre than an extract from non-overexpressing cells, but the pattern of telomerase addition products was indistinguishable in the two extracts.

Primer extension assays were carried out in telomerase buffer (20 mM Tris-

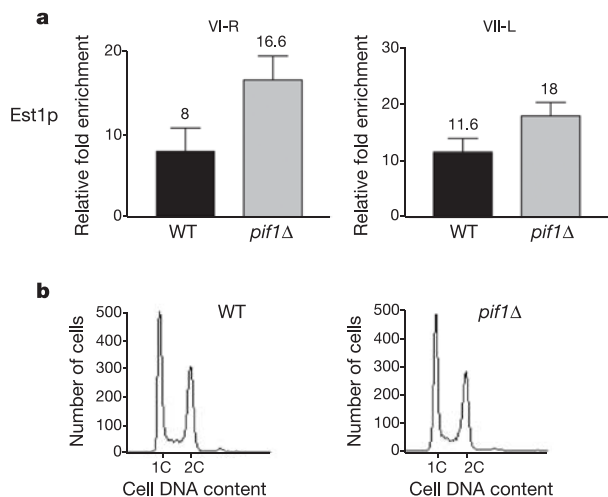


Figure 4 | Pif1p depletion increases telomere association of Est1p. *PIF1* (WT) or *pif1Δ* spore clones expressing Myc9-Est1p were grown to early log phase and analysed by ChIP and FACS. **a**, Quantification of telomere association of Myc9-Est1p with the VI-R (left) and VII-L (right) telomeres as in Fig. 3. When the same experiment was done for Est2p and Cdc13p, the amount of telomere association was not different in *PIF1* versus *pif1Δ* cells (see Supplementary Fig. S5). **b**, FACS analysis of cells for the ChIP analysis in **a**.

HCl at pH 8.0, 50 mM NaCl, 5 mM MgCl₂ and 1 mM spermidine) with 0.1 μ M of telomeric oligonucleotide TEL15 or TEL30. Wild-type Pif1p or Pif1p^{K264A} (1 μ M) was added when indicated. Reactions were started by addition of a nucleotide mix containing 50 μ M each of dGTP, dATP and dCTP, 5 μ M α -³²P-dTTP and 4 mM ATP, and incubated for 1 h. Reactions were stopped and processed as described²². Products were separated on 16% sequencing gels and quantified on a PhosphorImager.

Telomerase displacement assay. To monitor release of telomerase on Pif1p treatment, 1 μ M of a biotinylated Tel30 oligonucleotide bound to M-280 streptavidin dynabeads (Dyna) was added to 10 μ l of the telomerase extract in telomerase buffer and incubated for 10 min at room temperature. The beads were separated using a magnet, the supernatant was removed and the beads were resuspended in the same buffer with 4 mM ATP, 1 μ M Pif1p or 1 μ M Pif1p^{K264A} and 1 μ M oligonucleotide TEL15 (to trap the telomerase displaced from bead-bound oligonucleotides). Reactions were incubated at 30 °C for 10 min, and the beads again separated from the supernatant. Proteins attached to the oligonucleotides or in the supernatant were visualized by separation by 8% SDS polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting using anti-HAT polyclonal antiserum.

Chromatin immunoprecipitation. Chromatin immunoprecipitation experiments were carried out in YPH499 derived strains³⁵ that have a unique TEL sequence at the VII-L telomere and Myc-tagged alleles of *EST1*, *CDC13* or *EST2* (refs 3, 4, 36). Myc9-*EST1* and Myc9-*CDC13* have been described^{34,36}. *EST2* was tagged at the carboxy terminus such that the Myc epitopes were separated from the Est2p sequence by an 11-amino-acid flexible linker, similar to those previously described⁴. The linker used here had two repeats of three glycines and a serine, followed by three glycine residues. Each tagged protein was expressed from its native promoter at its endogenous locus and conferred to an approximate wild type telomere length⁴. Strains expressing epitope tagged proteins were transformed with the centromere plasmid pSH380 (ref. 37) containing no insert or the nuclear forms of either Pif1p or Pif1p^{K264A} under the control of the *GAL1* promoter. Strains were grown overnight in 3% raffinose, diluted to A₆₀₀ 0.055 in 3% raffinose, brought to 3% galactose, grown for 6 h and harvested for ChIP and FACS⁴. We used a relatively short pulse of Pif1p overexpression because extended exposure to high levels of wild-type or mutant Pif1p results in slow growth^{10,37}. Growth in galactose medium for 6 h had no significant effects on cell cycle progression, as monitored by FACS (Fig. 3c), nor on doubling time (data not shown). Each strain went through two population doublings in galactose medium. Using serial dilutions and western blotting, Pif1p and Pif1p^{K264A} were overexpressed to similar levels. The *pif1* Δ strains were obtained by sporulating diploids homozygous for the modified VII-L telomere and Myc-tagged protein and heterozygous at the *PIF1* locus that, before sporulation, were streaked 4–5 times to ensure wild-type telomere length in the starting diploids. Spores were dissected on glycerol plates, and individual spores used directly to inoculate YEPD cultures and grown for a total of ~30 cell divisions before analysis by ChIP and FACS. ChIP analysis, multiplex polymerase chain reaction (PCR) and quantification were as described⁴ except that immunoprecipitates were washed with 0.025% SDS, and 27 PCR cycles were used for the Cdc13p immunoprecipitates.

Received 24 March; accepted 3 August 2005.

Published online 24 August 2005.

1. Marcand, S., Brevet, V., Mann, C. & Gilson, E. Cell cycle restriction of telomere elongation. *Curr. Biol.* **10**, 487–490 (2000).
2. Diede, S. J. & Gottschling, D. E. Telomerase-mediated telomere addition *in vivo* requires DNA primase and DNA polymerases α and δ . *Cell* **99**, 723–733 (1999).
3. Taggart, A. K. P., Teng, S.-C. & Zakian, V. A. Est1p as a cell cycle-regulated activator of telomere-bound telomerase. *Science* **297**, 1023–1026 (2002).
4. Fisher, T. S., Taggart, A. K. P. & Zakian, V. A. Cell cycle-dependent regulation of yeast telomerase by Ku. *Nature Struct. Mol. Biol.* **11**, 1198–1205 (2004).
5. Schramke, V. et al. RPA regulates telomerase action by providing Est1p access to chromosome ends. *Nature Genet.* **36**, 46–54 (2004).
6. Schulz, V. P. & Zakian, V. A. The *Saccharomyces* Pif1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76**, 145–155 (1994).
7. Zhou, J.-Q., Monson, E. M., Teng, S.-C., Schulz, V. P. & Zakian, V. A. The Pif1p helicase, a catalytic inhibitor of telomerase lengthening of yeast telomeres. *Science* **289**, 771–774 (2000).
8. Mangahas, J. L., Alexander, M. K., Sandell, L. L. & Zakian, V. A. Repair of chromosome ends after telomere loss in *Saccharomyces*. *Mol. Biol. Cell* **12**, 4078–4089 (2001).
9. Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**, 1073–1076 (2001).
10. Lahaye, A., Stahl, H., Thines-Sempoux, D. & Foury, F. Pif1: a DNA helicase in yeast mitochondria. *EMBO J.* **10**, 997–1007 (1991).

11. Lue, N. F. Adding to the ends: what makes telomerase processive and how important is it? *Bioessays* **26**, 955–962 (2004).
12. Peng, Y., Mian, I. S. & Lue, N. F. Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance. *Mol. Cell* **7**, 1201–1211 (2001).
13. Collins, K. Ciliate telomerase biochemistry. *Annu. Rev. Biochem.* **68**, 187–218 (1999).
14. Greider, C. W. Telomerase is processive. *Mol. Cell. Biol.* **11**, 4572–4580 (1991).
15. Morin, G. B. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521–529 (1989).
16. Cohn, M. & Blackburn, E. H. Telomerase in yeast. *Science* **269**, 396–400 (1995).
17. Haering, C. H., Nakamura, T. M., Baumann, P. & Cech, T. R. Analysis of telomerase catalytic subunit mutants *in vivo* and *in vitro* in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **97**, 6367–6372 (2000).
18. Singh, S. M., Steinberg-Neifach, O., Mian, I. S. & Lue, N. F. Analysis of telomerase in *Candida albicans*: potential role in telomere end protection. *Eukaryot. Cell* **1**, 967–977 (2002).
19. Prowse, K. R., Avilion, A. A. & Greider, C. W. Identification of a nonprocessive telomerase activity from mouse cells. *Proc. Natl Acad. Sci. USA* **90**, 1493–1497 (1993).
20. Mantell, L. L. & Greider, C. W. Telomerase activity in germline and embryonic cells of *Xenopus*. *EMBO J.* **13**, 3211–3217 (1994).
21. Prescott, J. & Blackburn, E. Functionally interacting telomerase RNAs in the yeast telomerase complex. *Genes Dev.* **11**, 2790–2800 (1997).
22. Forstemann, K. & Lingner, J. Molecular basis for telomere repeat divergence in budding yeast. *Mol. Cell. Biol.* **21**, 7277–7286 (2001).
23. Jankowsky, E., Gross, C. H., Shuman, S. & Pyle, A. M. Active disruption of an RNA-protein interaction by a DEXH/D RNA helicase. *Science* **291**, 121–125 (2001).
24. Krejci, L. et al. DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
25. Veaute, X. et al. The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
26. Veaute, X. et al. UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *EMBO J.* **24**, 180–189 (2005).
27. Collins, K. & Greider, C. W. *Tetrahymena* telomerase catalyzes nucleolytic cleavage and nonprocessive elongation. *Genes Dev.* **7**, 1364–1376 (1993).
28. Hammond, P. W., Lively, T. N. & Cech, T. R. The anchor site of telomerase from *Euplotes aediculatus* revealed by photo-cross-linking to single- and double-stranded DNA primers. *Mol. Cell. Biol.* **17**, 296–308 (1997).
29. Shiratori, A. et al. Systematic identification, classification and characterization of the open reading frames which encode novel helicase-related proteins in *Saccharomyces cerevisiae* by gene disruption and northern analysis. *Yeast* **15**, 219–253 (1999).
30. Teixeira, M. T., Arneric, M., Sperisen, P. & Lingner, J. Telomere length homeostasis is achieved via a switch between telomerase-extendible and -nonextendible states. *Cell* **117**, 323–335 (2004).
31. Bessler, J. B., Torres, J. Z. & Zakian, V. A. The Pif1p subfamily of helicases: region specific DNA helicases. *Trends Cell Biol.* **11**, 60–65 (2001).
32. Zhou, J.-Q. et al. *Schizosaccharomyces pombe* pfh1+ encodes an essential 5' to 3' DNA helicase that is a member of the Pif1 sub-family of DNA helicases. *Mol. Biol. Cell* **13**, 2180–2191 (2002).
33. Bennett, R. J., Sharp, J. A. & Wang, J. C. Purification and characterization of the Sgs1 DNA helicase activity of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 9644–9650 (1998).
34. Burgers, P. M. Overexpression of multisubunit replication factors in yeast. *Methods* **18**, 349–355 (1999).
35. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
36. Tsukamoto, Y., Taggart, A. K. P. & Zakian, V. A. The role of the Mre11–Rad50–Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr. Biol.* **11**, 1328–1335 (2001).
37. Monson, E. K., Schulz, V. P. & Zakian, V. A. In *Genomic Instability and Immortality in Cancer* (eds Mihich, E. & Hartwell, L.) 97–110 (Plenum, New York, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank A. Chan for help with some of the experiments and J. Cooper, T. Fisher and M. Sabourin for comments on the manuscript. This work was supported by the NIH. J.B.B. was supported in part by a fellowship from the Association de la Recherche contre le Cancer and in part from a fellowship from the NJ Commission on Cancer Research; L.R.V. was supported in part by a Helen Hay Whitney post doctoral fellowship and in part by an NIH grant.

Author Contributions J.-B.B. and L.R.V. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.A.Z. (vzakian@molbio.princeton.edu).

LETTERS

A size of ~ 1 AU for the radio source Sgr A* at the centre of the Milky Way

Zhi-Qiang Shen¹, K. Y. Lo², M.-C. Liang³, Paul T. P. Ho^{4,5} & J.-H. Zhao⁴

Although it is widely accepted that most galaxies have super-massive black holes at their centres^{1–3}, concrete proof has proved elusive. Sagittarius A* (Sgr A*), an extremely compact radio source at the centre of our Galaxy, is the best candidate for proof^{5–7}, because it is the closest. Previous very-long-baseline interferometry observations (at 7 mm wavelength) reported that Sgr A* is ~ 2 astronomical units (AU) in size⁸, but this is still larger than the ‘shadow’ (a remarkably dim inner region encircled by a bright ring) that should arise from general relativistic effects near the event horizon of the black hole⁹. Moreover, the measured size is wavelength dependent¹⁰. Here we report a radio image of Sgr A* at a wavelength of 3.5 mm, demonstrating that its size is ~ 1 AU. When combined with the lower limit on its mass¹¹, the lower limit on the mass density is $6.5 \times 10^{21} M_{\odot} \text{pc}^{-3}$ (where $M_{\odot}$ is the solar mass), which provides strong evidence that Sgr A* is a super-massive black hole. The power-law relationship between wavelength and intrinsic size (size $\propto$ wavelength^{1.05}) explicitly rules out explanations other than those emission models with stratified structure, which predict a smaller emitting region observed at a shorter radio wavelength.

Past very-long-baseline interferometry (VLBI) observations^{12–16} of Sgr A* have revealed an east–west elongated structure whose apparent angular size at longer wavelengths is dominated by the interstellar scattering angle, that is, $\Theta_{\text{obs}} = \Theta_{\text{obs}}^{\text{I cm}} \lambda^2$, where Θ_{obs} is the observed size in milliarcseconds (mas) at wavelength λ in cm, and equals $\Theta_{\text{obs}}^{\text{I cm}}$ at 1 cm. Thus, VLBI observations at shorter millimetre wavelengths, where the intrinsic structure of Sgr A* could become comparable to the pure scattering size, are expected to show deviations of the observed size from the scattering law. This has been demonstrated by the recent detection of the intrinsic size at 7 mm (ref. 8). On 20 November 2002, we successfully carried out an observation of Sgr A* with the Very Long Baseline Array (VLBA) at its shortest wavelength of 3.5 mm (ref. 10). Our observation, with the steadily improved performance of the VLBA system, has produced the first (to our knowledge) high-resolution image of Sgr A* made at 3.5 mm (Fig. 1), which exhibits an elongated structure too.

To yield a quantitative description of the observed structure, we tried a model fitting procedure¹⁷ in which the amplitude closure relation is applied. Compared to the conventional VLBI

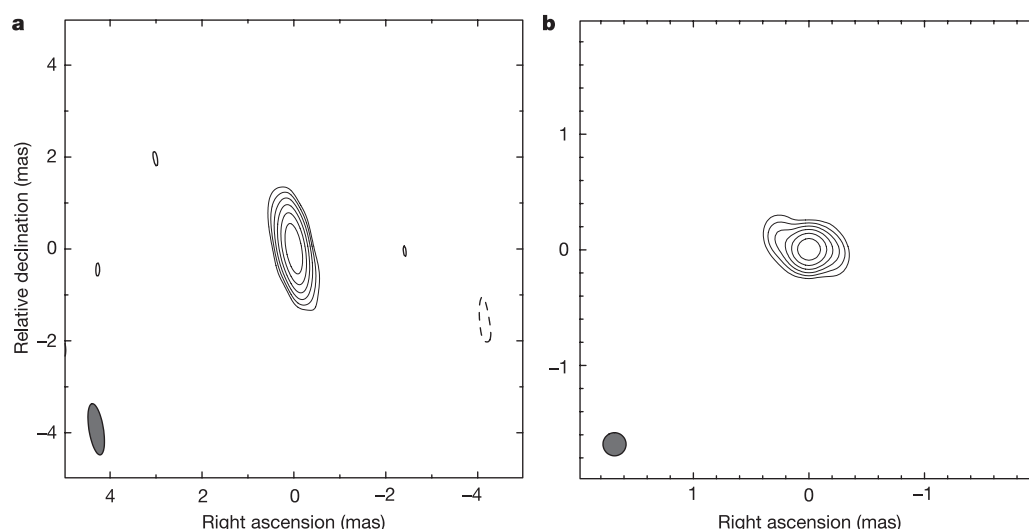


Figure 1 | High-resolution VLBI image of Sgr A* at 3.5 mm obtained with the VLBA on 20 November 2002. The observations were flexibly scheduled to ensure good weather conditions at most sites, and the data were recorded at the highest possible recording rate of 512 Mbit s^{−1}. Standard visibility amplitude calibration including the elevation-dependent opacity correction was done, and the final image was obtained after several iterations of the self-calibration and cleaning procedures. The calibrated total flux density is

about 1.2 Jy. **a**, A uniformly weighted image with the restoring beam (indicated at the lower left corner) of 1.13 mas $\times$ 0.32 mas at 9°. The peak flux density is 1.08 Jy beam^{−1}. Contour levels are drawn at $3\sigma \times (-1, 1, 2, 4, 8, 16, 32)$; $3\sigma = 17.5 \text{ mJy beam}^{-1}$. **b**, A super-resolution image with a circular beam of 0.20 mas from which an east–west elongated structure can be seen (see Table 1). Note the different scales. The contour levels are the same as that in **a** with the corresponding peak flux density of 1.01 Jy beam^{−1}.

¹Shanghai Astronomical Observatory, 80 Nandan Road, Shanghai 200030, China. ²National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903, USA. ³Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ⁴Harvard-Smithsonian CfA, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁵Institute of Astronomy & Astrophysics, Academia Sinica, PO Box 23-141, Taipei 106, Taiwan, China.

self-calibration and imaging technique, this can improve on the calibration of current VLBI observations of Sgr A* (see Supplementary Information). We have applied this model fitting procedure to 12 sets of VLBA observations of Sgr A* made at a variety of wavelengths from 6 cm to 3.5 mm over the time range from 1994 to 2004. Two experiments at 7 mm in 1994 are from the VLBA archive. Table 1 lists the fitting results. A consistent position angle ($\sim 80^\circ$) of the scatter-broadened image can be seen in all the data sets, regardless of the observing epoch and wavelength. Furthermore, it is quite significant that in all seven experiments made at 7 mm, the fitted apparent major axis size is always larger than the largest known scattering size of 0.69 mas extrapolated from the existing scattering models. We thus perform weighted least-squares fits to the near-simultaneous angular size measurements (in February 1997) as a function of the observing wavelength (see Supplementary Information for more details). We conclude that the best-fit two-dimensional scattering structure is $\Theta_{\text{major}} = (1.39 \pm 0.02)\lambda^2$ by $\Theta_{\text{minor}} = (0.69 \pm 0.06)\lambda^2$ with a position angle of $\sim 80^\circ$. Here Θ_{major} and Θ_{minor} are respectively the full-width at half-maximum of major and minor axes in mas. This gives an even smaller scattering angle along the major axis direction. An immediate important conclusion is that the discrepancy seen in all seven VLBI observations made at 7 mm is real, implying the appearance of the intrinsic source structure at wavelengths of 7 mm and shorter (see Supplementary Fig. 1).

At 7 mm, two measurements from the March 2004 observations are used to get averaged sizes of the major and minor axes of 0.724 ± 0.001 and 0.384 ± 0.013 mas, respectively, with position angle $80.6^{+0.5}_{-0.6}$ degrees. The difference between the measured and the extrapolated scattering sizes along the major axis is $\Delta\theta = 0.053 \pm 0.010$ mas, significant at $>5\sigma$ level. By subtracting in quadrature the scattering angle, this suggests an intrinsic size of 0.268 ± 0.025 mas for the major axis. Similarly, we can derive an intrinsic size for the minor axis to be 0.190 ± 0.057 mas, comparable to the major axis size. However, it should be kept in mind that the deviation seen for the minor axis is only significant at the 1.6σ level. Note that the derived source size has greater statistical significance

than the deviation, because the scattering size has been deduced with good accuracy.

At 3.5 mm, the fitted apparent source structure from November 2002 VLBA observations is $0.21^{+0.02}_{-0.01}$ mas by $0.13^{+0.05}_{-0.13}$ mas, with a position angle of 79^{+12}_{-33} degrees (see Table 1). Thus, an intrinsic size of 0.126 ± 0.017 mas for the major axis can be obtained from the deviation $\Delta\theta = 0.042 \pm 0.010$ mas (at $>4\sigma$ level) in the measured major axis size from that of the scattering angle at 3.5 mm. Mainly owing to the limited resolution, the minor axis measurement at 3.5 mm is, however, inadequate to make any firm claim on the determination of its intrinsic size. So, we defer any estimate of intrinsic minor axis size at 3.5 mm for future investigation. Past 3.5 mm VLBI observations with the heterogeneous Coordinated Millimeter VLBI Array, severely limited by its low sensitivity, could not warrant a model more complex than the circular one¹⁸. The best-fit circular gaussian has a diameter of 0.18 ± 0.02 mas, which is indistinguishable from the scattering size along the major axis and cannot give a meaningful estimate of the intrinsic structure⁸.

Thus we have sampled a zone of the supermassive black hole (SMBH) closer to the event horizon than ever before, by detecting the intrinsic size of Sgr A* to be only 1.01 AU at a distance of 8.0 kpc, or $12.6R_{\text{sc}}$, where $R_{\text{sc}} (= 1.2 \times 10^{12} \text{ cm})$ is the Schwarzschild radius of a $4 \times 10^6 M_\odot$ SMBH. By assuming a spherical structure, we obtain a lower limit to the mass density of Sgr A* of $6.5 \times 10^{21} M_\odot \text{ pc}^{-3}$. Here, we used the lower bound to the mass of Sgr A*, derived from the upper limit to the intrinsic proper motion of Sgr A* itself¹, which is about 10% of the $4 \times 10^6 M_\odot$ inferred from the stellar orbital motions^{6–7}, and the upper limit to the source intrinsic size (this work). This mass density is at least four orders of magnitude greater than that determined from dynamical measurements of stellar velocities⁶. This is because here we are probing directly the structure of Sgr A*, where the assumed mass estimate refers to the value within Sgr A*. We note also that this mass density is almost 12 orders of magnitude greater than the estimate for NGC 4258 (ref. 19), one of the best known SMBHs. Such an extraordinarily high mass density robustly rules out the possibility of Sgr A* being a compact dark cluster of stellar remnants as it would have an unreasonably short lifetime of less than 100 yr (ref. 20), and thus argues strongly in favour of the SMBH nature of Sgr A*. To prove that Sgr A* is indeed an SMBH requires an unambiguous demonstration that Sgr A* possesses an event horizon. It is intriguing that the detected intrinsic size at 3.5 mm is about twice the diameter of the shadow caused by

Table 1 | Parameters of the elliptical gaussian model for Sgr A*

λ (cm)	ν , bw, rate† (GHz) (MHz) (bits)	Epoch (d month yr)	$\Theta_{\text{major}} \ddagger$ (mas)	$\Theta_{\text{minor}} \ddagger$ (mas)	PA§ (°)
0.348	86.236, 128, 2	03 Nov. 2002	$0.21^{+0.02}_{-0.01}$	$0.13^{+0.05}_{-0.13}$	79^{+12}_{-33}
0.694	43.213, 32, 2	14 Feb. 1997	$0.71^{+0.01}_{-0.01}$	$0.42^{+0.05}_{-0.05}$	74^{+2}_{-2}
0.695	43.175, 128, 2	08 Mar. 2004	$0.722^{+0.002}_{-0.002}$	$0.395^{+0.019}_{-0.020}$	$80.4^{+0.8}_{-0.8}$
0.695	43.175, 128, 2	20 Mar. 2004	$0.725^{+0.002}_{-0.002}$	$0.372^{+0.020}_{-0.018}$	$80.8^{+0.6}_{-0.9}$
0.695	43.151, 64, 1	26 Apr. 1994	$0.72^{+0.01}_{-0.01}$	$0.39^{+0.07}_{-0.07}$	78^{+2}_{-2}
0.695	43.151, 64, 1	29 Sep. 1994	$0.72^{+0.01}_{-0.01}$	$0.42^{+0.03}_{-0.03}$	79^{+1}_{-1}
0.695	43.135, 32, 2	24 Apr. 1999	$0.69^{+0.01}_{-0.01}$	$0.33^{+0.04}_{-0.04}$	83^{+1}_{-1}
0.695	43.135, 32, 2	23 May 1999	$0.71^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$	79^{+1}_{-1}
1.350	22.229, 32, 1	12 Feb. 1997	$2.53^{+0.06}_{-0.05}$	$1.45^{+0.23}_{-0.38}$	83^{+4}_{-5}
1.953	15.361, 32, 1	12 Feb. 1997	$5.33^{+0.07}_{-0.07}$	$2.70^{+0.30}_{-0.44}$	83^{+3}_{-3}
3.564	8.417, 32, 1	07 Feb. 1997	$17.5^{+0.5}_{-1.0}$	$8.5^{+1.0}_{-1.0}$	87^{+3}_{-3}
6.020	4.983, 32, 1	07 Feb. 1997	$43.0^{+2.5}_{-1.0}$		

Results from the model fitting procedure, which implicitly uses the amplitude closure relation (see Supplementary Information for details). Also listed are some details of the VLBI observations. Except for five observations in February 1997 which used the VLBA and one VLA antenna, all the other observations were performed by the VLBA. λ , wavelength; ν , frequency; Θ_{major} , Θ_{minor} and PA, parameters of gaussian model: respectively major and minor axes, and the position angle of the major axis.

† The observing frequency in GHz, the recording bandwidth in MHz, and the sampling rate (1 bit or 2 bits). With the same total bandwidth, the noise level from 2-bit quantization data is about $\sqrt{2}$ times lower than that from 1-bit quantization. The recording rate is two times the product of the bandwidth and the sampling rate.

‡ Full-width at half-maximum (FWHM).

§ In all cases, errors are 1σ .

|| Dual polarization observation. Both left and right circular polarization data have the same recording mode (listed), and both data were used in the model fitting.

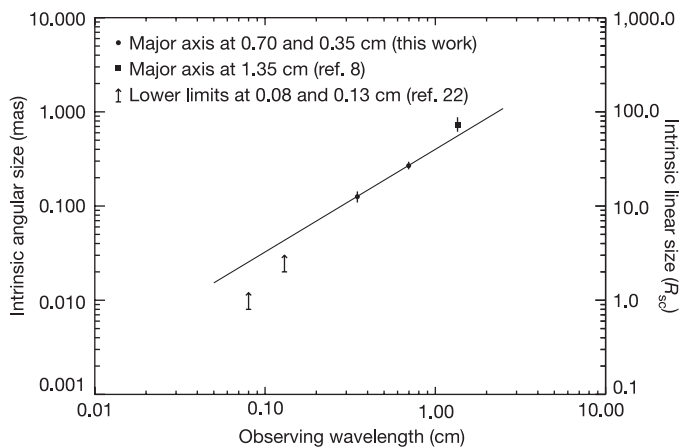


Figure 2 | Intrinsic major axis size versus observing wavelength. The solid line represents the two-point fit from the detected intrinsic sizes at both 3.5 and 7 mm (this work). Also plotted are the lower limits to the intrinsic sizes at 1.3 and 0.8 mm (ref. 22), and the reported detection at 1.35 cm (ref. 8). The extrapolated intrinsic sizes at 1.3 and 0.8 mm are about $4.3R_{\text{sc}}$ and $2.5R_{\text{sc}}$ respectively. Here $R_{\text{sc}} (= 1.2 \times 10^{12} \text{ cm})$ is the Schwarzschild radius for a $4 \times 10^6 M_\odot$ SMBH. The error bars represent 1σ standard deviation.

the strong gravitational bending of light rays⁹. Thus, it is very likely that VLBI observations of Sgr A* at 1 mm or shorter will reach the region comparable to its shadow, which can be used to differentiate between the SMBH scenario and other supermassive non-baryonic stars^{9,21}.

The two-point fit to the well determined intrinsic sizes at 7 and 3.5 mm shows a λ^β -dependence of the intrinsic source size with $\beta = 1.09^{+0.34}_{-0.32}$ (Fig. 2). Also plotted in Fig. 2 are two inferred lower limits of 0.02 and 0.008 mas to the intrinsic size at 1.3 and 0.8 mm, respectively, from the absence of refractive scintillation²². These lower limits are consistent with the extrapolation of the λ^β -dependence. However, we note that these two lower limits are only about 2 and 0.8 R_{sc} , which are smaller than the last stable orbit (LSO) radius of $3R_{\text{sc}}$ for a non-rotating (Schwarzschild) black hole. There is some evidence that Sgr A* is a rotating black hole²³. For a prograde maximally rotating Kerr black hole, the LSO radius is $0.5R_{\text{sc}}$. The LSO establishes the lower limit to the emission region size. Hence the λ^β -dependence will eventually reach a minimum. As such, the turn-over frequency²⁴ seen in the entire spectrum of Sgr A* might tell us the smallest size of the emission, which can be further used to constrain its spin, if any.

The extrapolated intrinsic size at 1.35 cm is $0.555^{+0.136}_{-0.115}$ mas. This, when compared to the scattering angle of 2.576 ± 0.036 mas, is consistent with the idea that the scattering effect dominates the observed source size at 1.35 cm. This deduced source size is also formally consistent with the reported detection of $0.726^{+0.152}_{-0.111}$ mas, to within the uncertainties⁸.

The derived $\lambda^{1.09}$ -dependence requires that the emission at different wavelengths is dominated by different emitting regions, and thus conclusively excludes those models without the stratified emission structure. Along with the detected intrinsic major axis size, we can derive a lower limit to the intrinsic brightness temperature as $T_b \geq 1.36 \times 10^9 \times \frac{S_\lambda \lambda^2}{\theta_{\text{int}}^2}$ K; here S_λ is the flux density in Jy at wavelength λ in cm, and θ_{int} is the intrinsic major axis size in mas. There is a wavelength dependence of the lower limit T_b as $\lambda^{-\alpha-0.18}$ (assuming $S_\lambda \propto \lambda^{-\alpha}$). Using the flux densities of 1.0 and 1.2 Jy at 7 and 3.5 mm, respectively, the corresponding minimal T_b is 0.9×10^{10} and 1.2×10^{10} K, greater than the prediction of the spherical accretion model²⁵. However, this lower limit of 10^{10} K, and the spatial distribution of the radio emission, can be explained easily by the inhomogeneous jet model^{26,27}, in which the magnetic field and the electron number density vary with the distance (r) to the origin of the jet as r^{-1} and r^{-2} , respectively. On the other hand, the radiatively inefficient accretion flow (RIAF) model of Sgr A* (ref. 28) can also account for a brightness temperature of $>10^{10}$ K as well as the observed spectral energy distribution. The prediction of $\lambda^{0.9}$ from the hybrid thermal-nonthermal synchrotron radiation from RIAF (ref. 29) is in agreement with the estimated $\lambda^{1.09+0.34}_{-0.32}$ relation. Here, the possible existence of strong outflows from the accretion disk was not taken into account. To further discern between them, it is important to study the correlation between the detected X-ray variability and the variations frequently seen in the radio to sub-millimetre wavelengths, which would yield further information on the intrinsic density structure of the emitting zone.

Received 1 July; accepted 31 August 2005.

1. Begelman, M. C. Evidence for black holes. *Science* **300**, 1898–1903 (2003).
2. Kormendy, J. & Richstone, D. Inward bound—The search for supermassive black

- holes in galactic nuclei. *Annu. Rev. Astron. Astrophys.* **33**, 581–624 (1995).
3. Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
4. Balick, B. & Brown, R. L. Intense sub-arcsecond structure in the Galactic Center. *Astrophys. J.* **194**, 265–270 (1974).
5. Melia, F. & Falcke, H. The supermassive black hole at the Galactic Center. *Annu. Rev. Astron. Astrophys.* **39**, 309–352 (2001).
6. Schödel, R. *et al.* A star in a 15.2 year orbit around the supermassive black hole at the centre of the Milky Way. *Nature* **419**, 694–696 (2002).
7. Ghez, A. M. *et al.* Stellar orbits around the Galactic Center black hole. *Astrophys. J.* **620**, 744–757 (2005).
8. Bower, G. C. *et al.* Detection of the intrinsic size of Sagittarius A* through closure amplitude imaging. *Science* **304**, 704–708 (2004).
9. Falcke, H., Melia, F. & Agol, E. Viewing the shadow of the black hole at the Galactic Center. *Astrophys. J.* **528**, L13–L16 (2000).
10. Shen, Z.-Q. & Lo, K. Y. High-resolution 86 GHz VLBA imaging of Sgr A*. *Prog. Theor. Phys. Suppl.* **155**, 413–414 (2004).
11. Reid, M. *et al.* The position, motion, and mass of Sgr A*. *Astron. Nachr.* **324** (Suppl. Iss. 1), 505–511 (2003).
12. Davies, R. D., Walsh, D. & Booth, R. S. The radio source at the Galactic nucleus. *Mon. Not. R. Astron. Soc.* **177**, 319–333 (1976).
13. Lo, K. Y. *et al.* On the size of the galactic centre compact radio source: diameter <20 AU. *Nature* **315**, 124–126 (1985).
14. Alberdi, A. *et al.* VLBA Image of Sgr A* at $\lambda = 1.35$ cm. *Astron. Astrophys.* **277**, L1–L4 (1993).
15. Bower, G. C. & Backer, D. C. 7 Millimeter VLBA observations of Sagittarius A*. *Astrophys. J.* **496**, L97–100 (1998).
16. Lo, K. Y., Shen, Z.-Q., Zhao, J.-H. & Ho, P. T. P. Intrinsic size of Sagittarius A*: 72 Schwarzschild radii. *Astrophys. J.* **508**, L61–L64 (1998).
17. Shen, Z.-Q., Liang, M. C., Lo, K. Y. & Miyoshi, M. Searching for structural variability in Sgr A*. *Astron. Nachr.* **324** (Suppl. Iss. 1), 383–389 (2003).
18. Doeleman, S. S. *et al.* Structure of Sagittarius A* at 86 GHz using VLBI closure quantities. *Astron. J.* **121**, 2610–2617 (2001).
19. Miyoshi, M. *et al.* Evidence for a black hole from high rotation velocities in a sub-parsec region of NGC4258. *Nature* **373**, 127–129 (1995).
20. Maoz, E. Dynamical constraints on alternatives to supermassive black holes in galactic nuclei. *Astrophys. J.* **494**, L181–L184 (1998).
21. Torres, D. F., Capozziello, S. & Lambiase, G. Supermassive boson star at the galactic center? *Phys. Rev. D* **62**, 104012 (2000).
22. Gwinn, C. R., Danen, R. M., Tran, T. Kh., Middleditch, J. & Ozerov, L. M. The Galactic center radio source shines below the Compton limit. *Astrophys. J.* **381**, L43–L46 (1991).
23. Genzel, R. *et al.* Near-infrared flares from accreting gas around the supermassive black hole at the Galactic Centre. *Nature* **425**, 934–937 (2003).
24. Falcke, H. *et al.* The simultaneous spectrum of Sagittarius A* from 20 centimeter to 1 millimeter and the nature of the millimeter excess. *Astrophys. J.* **499**, 731–734 (1998).
25. Melia, F. An accretion black hole model for Sagittarius A*. II. A detailed study. *Astrophys. J.* **426**, 577–585 (1994).
26. Königl, A. Relativistic jets as X-ray and gamma-ray sources. *Astrophys. J.* **243**, 700–709 (1981).
27. Falcke, H. & Markoff, S. The jet model for Sgr A*: Radio and X-ray spectrum. *Astron. Astrophys.* **362**, 113–118 (2000).
28. Yuan, F., Quataert, E. & Narayan, R. Nonthermal electrons in radiatively inefficient accretion flow models of Sagittarius A*. *Astrophys. J.* **598**, 301–312 (2003).
29. Özel, F., Psaltis, D. & Narayan, R. Hybrid thermal-nonthermal synchrotron emission from hot accretion flows. *Astrophys. J.* **541**, 234–249 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Very Large Array and the Very Long Baseline Array are operated by the National Radio Astronomy Observatory, which is a facility of the National Science Foundation, operated under cooperative agreement by Associated Universities Inc. Z.-Q.S. acknowledges support from the One-Hundred-Talent programme of the Chinese Academy of Sciences

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.-Q.S. (zshen@shao.ac.cn).

Active control of slow light on a chip with photonic crystal waveguides

Yurii A. Vlasov¹, Martin O'Boyle¹, Hendrik F. Hamann¹ & Sharee J. McNab¹

It is known that light can be slowed down in dispersive materials near resonances¹. Dramatic reduction of the light group velocity—and even bringing light pulses to a complete halt—has been demonstrated recently in various atomic^{2–5} and solid state systems^{6–8}, where the material absorption is cancelled via quantum optical coherent effects^{3–5,7}. Exploitation of slow light phenomena has potential for applications ranging from all-optical storage to all-optical switching^{9,10}. Existing schemes, however, are restricted to the narrow frequency range of the material resonance, which limits the operation frequency, maximum data rate and storage capacity¹⁰. Moreover, the implementation of external lasers, low pressures and/or low temperatures prevents miniaturization and hinders practical applications. Here we experimentally demonstrate an over 300-fold reduction of the group velocity on a silicon chip via an ultra-compact photonic integrated circuit using low-loss silicon photonic crystal waveguides^{11,12} that can support an optical mode with a submicrometre cross-section^{13,14}. In addition, we show fast (~ 100 ns) and efficient (2 mW electric power) active control of the group velocity by localized heating of the photonic crystal waveguide with an integrated micro-heater.

The propagation of light pulses in a dielectric medium with refractive index $n(\lambda)$ is described by the group velocity, defined as $V_g = c/(n - \lambda dn/d\lambda) = c/n_g$, where c is the speed of light in vacuum, n is the phase refractive index, λ is the wavelength, and n_g is the group index. When dispersion is negative and large ($dn/d\lambda < 0$), the pulses can be significantly delayed with respect to free space propagation. Slow group velocity was measured recently in photonic crystal structures with ultrafast pulse propagation techniques^{15,16}. Surface coupling to slow light modes was also inferred from the absence of transmission of light at wavelengths corresponding to strong dispersion^{17,18}. However, the accuracy of the group velocity determination in both of these approaches is limited, as they are relying heavily on measurements of the amplitude of the transmitted light, whereas the dispersion is inherently connected with its phase. For example, significant amplitude reshaping of short optical pulses owing to strong group velocity dispersion makes the group delay assignment progressively inaccurate¹⁵ in the slow group velocity regime. In contrast, phase-sensitive optical techniques based on observation of interference fringes in transmission spectra were successfully used to accurately measure group indices approaching 100 (refs 14, 19). To utilize this interferometric approach, we designed and fabricated an integrated Mach–Zehnder interferometer (MZI) employing photonic crystal waveguides.

Silicon photonic crystal waveguides used in our experiments are shown in Fig. 1. These are resonant photonic structures formed by etching a periodic array (periodicity $a = 437$ nm) of holes with radius $0.25a$ in a 223-nm-thick silicon suspended membrane. The light is coupled to the photonic crystal waveguide through a polymer-based fibre coupler and single-mode access strip waveguide butt-coupled to the photonic crystal¹¹. Details of the structural parameters

and device fabrication are described in Methods. The utilization of laterally tapered spot-sized converters and careful termination of the photonic crystal lattice¹⁹ at a position half way through the holes nearest to the waveguide (see Fig. 1a) allows efficient coupling to the

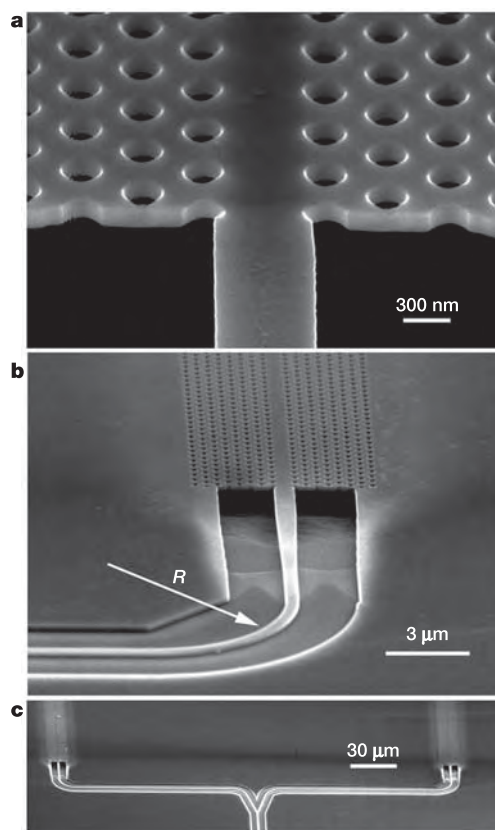


Figure 1 | SEM images of a passive unbalanced Mach–Zehnder interferometer using photonic crystal waveguides. **a**, Input section of the photonic crystal waveguide showing the suspended silicon membrane etched with holes and butt-coupled to a strip waveguide. The termination of the photonic crystal lattice at the coupling interface is chosen to obtain highest coupling efficiency in the slow light regime. **b**, Broader view of the photonic crystal waveguide membrane and input strip waveguide. After passing through a sharp 90° bend with radius $R = 5 \mu\text{m}$, the mode is widened in the tapered section to better match the photonic crystal slow light mode. **c**, View of the input of the Mach–Zehnder interferometer (MZI) with reference (left) and signal (right) arms and a compact 15° angle Y-junction that splits the light equally between the arms. The output side of the optical circuit has an analogous Y-junction and is terminated by a single output strip waveguide.

¹IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA.

slow light mode. A typical transmission spectrum of a straight photonic crystal waveguide of $250\text{ }\mu\text{m}$ length is shown in Fig. 2a (black curve). According to photonic band structure calculations performed with a plane wave expansion method²⁰ (black dashed curve in Fig. 2b, see also Methods), the group index becomes increasingly large within a 20 nm bandwidth around $1,620\text{ nm}$ in the vicinity of the sharp cut-off of the TE-like waveguiding mode.

In order to measure the group velocity interferometrically, an integrated unbalanced MZI with a footprint of only 0.04 mm^2 was designed (Fig. 1c). The light is divided equally at the compact 15° -angle Y-splitter and is passed through two almost identical photonic crystal waveguides (Fig. 1c) with a small but noticeable difference in the hole radii of $0.251a$ (reference arm) and $0.263a$ (signal arm). Increasing the hole radii in the signal arm shifts the photonic crystal mode cut-off by approximately 20 nm to shorter wavelengths at around $1,600\text{ nm}$. In the spectral region of interest (between $1,500\text{ nm}$ and $1,600\text{ nm}$), the group index in the reference arm can be considered almost constant, $n_g^{\text{ref}} \approx 5$ (see dashed black curve in Fig. 2b), while the group index of the signal arm $n_g^{\text{sig}}(\lambda)$ exhibits a strong increase. Correspondingly, when the light from both arms is recombined at the output Y-coupler of the MZI, strong cosine-like interference fringes with a modulation depth exceeding 25 dB are observed in the transmission spectrum (red curve in Fig. 2a). The fringes converge towards the cut-off of the photonic crystal mode in the signal arm at around $1,600\text{ nm}$. The spectral distance between corresponding minima and maxima significantly decreases from 20 nm at $1,510\text{ nm}$ to numbers of the order of 0.01 nm around $1,602.8\text{ nm}$, reflecting a sharp increase in the relative phase shift. Approaching the cut-off of the signal arm, the fringe visibility decreases sharply from $\sim 25\text{ dB}$ to $\sim 2\text{ dB}$, indicating significant wavelength dependent losses in the circuit. The transmission above $1,602.8\text{ nm}$ is provided only by the reference arm up to the cut-off wavelength of that arm around $1,620\text{ nm}$.

The maxima and minima of the oscillations correspond to constructive and destructive interference, respectively, with the relative phase shift between two adjacent extrema equal to π . The spectral dependence of the group index in the signal arm can then be deduced from positions of corresponding minima λ_{min} and maxima λ_{max} as $n_g^{\text{sig}}(\lambda) = \lambda_{\text{min}}\lambda_{\text{max}}/[2L(\lambda_{\text{min}} - \lambda_{\text{max}})] + n_g^{\text{ref}}(\lambda)$, where $L = 250\text{ }\mu\text{m}$ is the length of the photonic crystal waveguide. The slight wavelength dependence of the group index $n_g^{\text{ref}}(\lambda)$ in the reference arm was taken into account using exponential fitting of the theoretical black dashed curve in Fig. 2b as $n_g^{\text{ref}}(\lambda) = 1.2 \times 10^{-32} \exp(\lambda/21.32) + 4.06$. Blue circles in Fig. 2b represent the group index deduced from the transmission spectrum as described above, and exhibit group indices in excess of 300. The fringe visibility decreases towards the signal arm cut-off, which makes the assignment of transmission peaks to MZI-related interference fringes difficult. Detailed discussion of the limitations of the method is presented in Supplementary Information. In summary, the fringe amplitude and fringe spacing at these wavelengths becomes comparable with that of a superimposed residual Fabry–Perot (FP) noise with a similar amplitude ($\sim 1.5\text{ dB}$) that arises from interference in the optical circuit outside the MZI. Fortunately, the main oscillation period of the FP noise ($\sim 0.15\text{ nm}$) is almost constant throughout the spectrum, which corresponds to the distance of $\sim 1\text{ mm}$ between the MZI and the input fibre couplers at the facets of the photonic chip. This allows even small fringes at wavelengths longer than $1,602\text{ nm}$ (shown in Fig. 2a inset) to be assigned as arising from the MZI. The last few data points around $1,602.65\text{ nm}$ correspond to very large group indices of over 500. However, the experimental error here is also large, since the fringe amplitude is approaching that of FP noise. The group index spectral dependence deviates from a smooth monotonically increasing theoretical curve calculated for infinitely long structures, and in turn also exhibits oscillations converging towards the photonic crystal mode cut-off (Fig. 2b inset). These oscillations can be readily attributed to

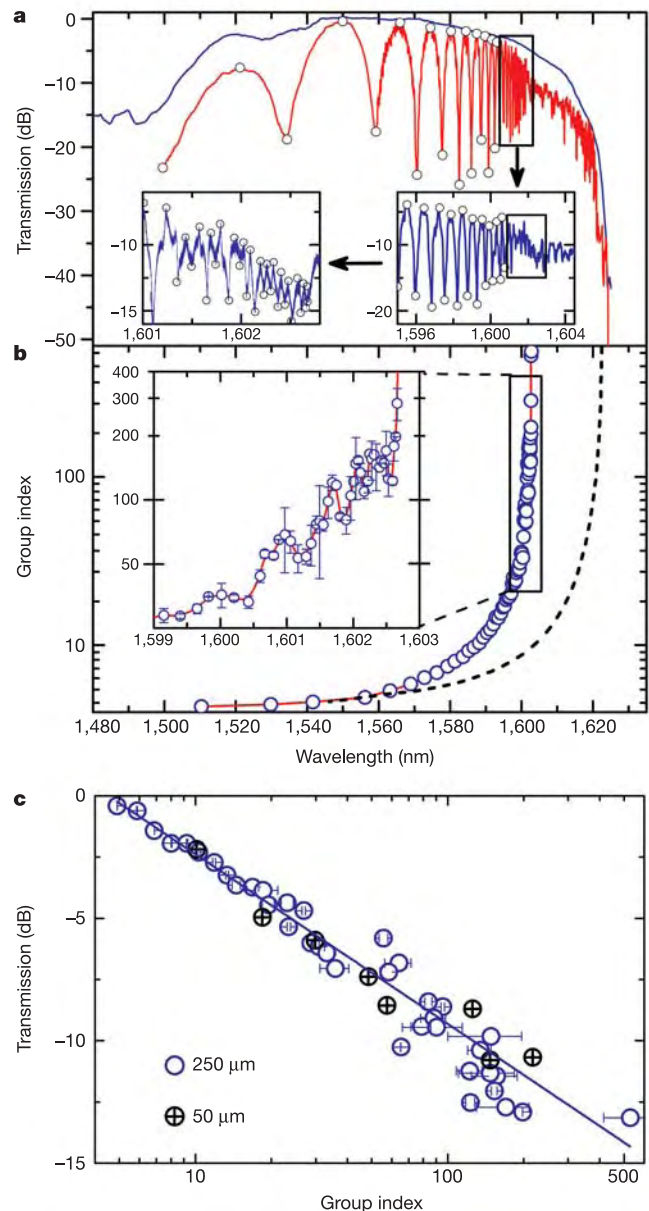


Figure 2 | Optical measurements of a passive unbalanced MZI.

a, Transmission spectra of a straight photonic crystal waveguide with hole radii $0.251a$ and length $250\text{ }\mu\text{m}$ (black curve) and of a MZI (red curve) with $250\text{-}\mu\text{m}$ -long photonic crystal waveguides having hole radii $0.251a$ in the reference arm and $0.263a$ in the signal arm. Open blue circles represent spectral positions of the minima and maxima of the interference fringes used for calculation of the group index. The two insets show a consecutively magnified view of the spectrum near the transmission cut-off of the signal arm. **b**, Semi-logarithmic plot of the group index in the signal arm (open blue circles) extracted from the transmission spectrum. Inset shows the blown-up portion of the main plot near the transmission cut-off of the signal arm. Dashed black line shows the results of three-dimensional plane wave calculations of the group index in the reference arm. **c**, Amplitude of interference fringe maxima as a function of the group index for two MZI with photonic crystal waveguide lengths of $50\text{ }\mu\text{m}$ (crossed black circles) and $250\text{ }\mu\text{m}$ (open blue circles). Blue curve is a fitting with a power law dependence $T = a(n_g)^b$, where $a = 3.22$ and $b = -0.74$. The procedure for error determination is described in the Supplementary Discussion. The error bars are estimated from the maximum value of uncertainty in defining the spectral position of each individual extremum in transmission spectra.

Bloch–Floquet modes (or group delay ripples) in the finite-length photonic crystal waveguide²¹.

To determine the potential of photonic crystal waveguides with large group indices in applications such as all-optical buffering, the chromatic dispersion and propagation losses need to be analysed. A strong increase of the group index close to the cut-off wavelength results in a sharp increase of the group velocity dispersion. This chromatic dispersion can be estimated from the results of Fig. 2b as $500 \text{ ps nm}^{-1} \text{ mm}^{-1}$ for group indices around 100. Compared with the chromatic dispersion in conventional telecommunications fibres, this is approximately 10^7 times higher. This induces significant broadening of ultrafast optical pulses transmitted through a photonic crystal waveguide, as observed in recent experiments¹⁶. Simple estimations show, however, that this extremely large dispersion might still allow operation of an all-optical buffer incorporating a millimetre-long photonic crystal waveguide that can store three bits of information at 10 Gbit s^{-1} data rate (ref. 10).

Propagation and coupling losses in the slow light regime can in principle result in the loss of stored information. The total loss in the photonic crystal waveguide as a function of the group index is analysed in Fig. 2c, where the amplitudes of the fringe maxima are shown (blue circles). Comparison with the results from another MZI device with five times shorter photonic crystal waveguides ($L = 50 \mu\text{m}$ in each arm; black crossed circles in Fig. 2c) shows that propagation losses are virtually absent and that the fringe visibility is mainly affected by increasingly large coupling losses, which is in agreement with previously published data¹⁹. Thus further engineering of the coupling interface is essential to promote efficient coupling of light into the photonic crystal mode if even higher group indices (longer group delays) are to be accessed.

Slow light propagation can be exploited for a broader range of applications^{9,10}—for example, variable optical buffers or dynamic dispersion compensators—if it can be tuned by external electrical signals. This can be done, for example, by passing an electric current

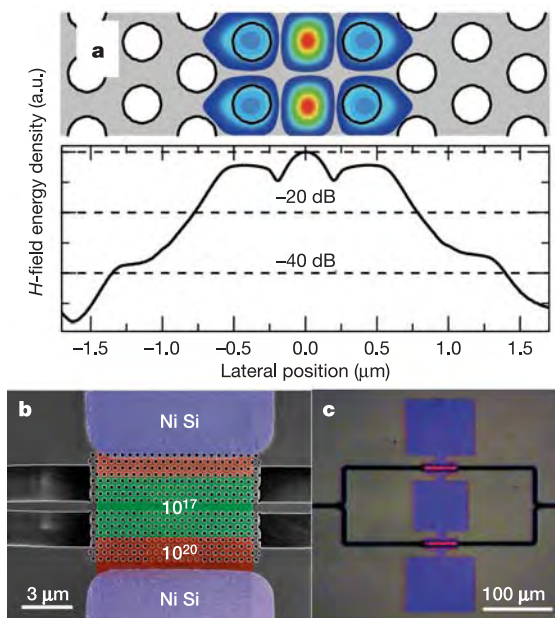


Figure 3 | Active electrically tunable MZI with lateral electrical contacts to photonic crystal waveguides. **a**, Time averaged magnetic field energy density calculated by the plane wave method in top-down view (false colour linear scale) and plotted on a semi-logarithmic scale (integrated over one unit cell). **b**, SEM image of the fabricated photonic crystal waveguide with lateral electrical contacts superimposed with a false colour representation to denote the different doping levels. **c**, Optical micrograph of the completed MZI circuit with three electrical contacts.

between metallic contacts and thereby locally heating the signal arm of the MZI^{22–25}. However, electrical contacts placed directly onto the silicon waveguide in the immediate vicinity of the strongly confined photonic crystal mode result in large absorption losses^{22,23}. Increasing the distance between the photonic crystal mode and the contacts helps to minimize absorption, but requires larger electrical powers to change the refractive index^{22,24,25}. Here we present a solution to this problem based on the idea of lateral electrical contacts to the photonic crystal, which itself acts as a micro-heater^{26,27}. Figure 3a shows the magnetic field energy density profile of the photonic crystal mode calculated by the plane wave expansion method²⁰ for the slow light regime, showing that over 90% of the mode energy is confined within only 500 nm from the waveguide centre. Excellent optical isolation exceeding -50 dB is obtained with only four rows of holes on each side of the waveguide. Implementing this idea, ohmic lateral electrical contacts were defined on top of the photonic crystal

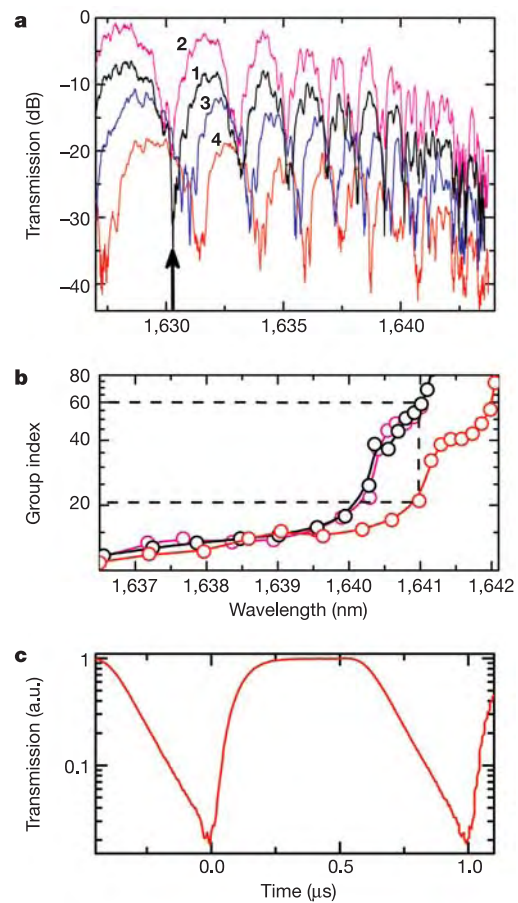


Figure 4 | Thermo-optic tuning of the group index in the active unbalanced MZI. **a**, Set of transmission spectra measured at the output of the MZI for different applied electric power. Spectra are shifted vertically by 5 dB for clarity. Spectrum 1 (black) corresponds to zero applied power. Spectrum 2 (magenta) was measured with 1.5 mW power applied to the reference arm, and is almost identical to spectrum 1 except for a slight shift of the fringes to shorter wavelengths. Spectrum 3 (blue) and spectrum 4 (red) were measured with 1 mW and 2 mW electric power, respectively, applied to the signal arm. A large shift of the fringes to longer wavelengths is evident. **b**, Semi-logarithmic plot of the group index dispersion extracted from the spectra in **a**. Black circles correspond to spectrum 1 with zero applied power. Magenta circles correspond to spectrum 3 with 1.5 mW applied to the reference arm. Red circles correspond to spectrum 4 with 2 mW power applied to the signal arm. **c**, Time response of the MZI with a 1 MHz square-wave modulated electric signal corresponding to 2 mW electrical power applied to the signal arm. The wavelength of the laser is tuned to 1,630 nm, as shown by the arrow in **a**.

waveguide separated by $1.5\ \mu\text{m}$ from the waveguide core, as shown in Fig. 3b. Details of the design and fabrication are presented in Methods and Supplementary Information. Three contacts to the MZI device were formed as shown in Fig. 3c, which allows both the signal and reference arms to be addressed independently.

Transmission spectra for an unbalanced active MZI are presented in Fig. 4a. Owing to different structural parameters of the active MZI, the cut-offs of the signal and reference arms occur at around $1,642\ \text{nm}$ and $1,655\ \text{nm}$, respectively. Measurements are shown for different electric power levels applied consecutively to the signal and reference arms. Using the procedure described above, the group index of the signal arm can be deduced from the spectral positions of the interference fringes. Figure 4b represents three typical measurements of the group indices of the signal arm with and without applied electric power. Local heating of the signal arm results in the wavelength shift of the group index dispersion curve to longer wavelengths. For a given wavelength, for example $1,641\ \text{nm}$, the group index is tuned from 20 to 60 with $2\ \text{mW}$ of electric power applied to the signal arm (red curve in Fig. 4b). An even larger tuning range can be achieved at longer wavelengths. But because the active MZI device underwent nine additional fabrication steps (see Supplementary Information), the amplitude of residual FP noise is increased, which prevents accurate measurements of group indices higher than 80. As expected, heating of the reference arm (with its group index nearly constant throughout this spectral region) does not result in significant changes of the measured group index dispersion (magenta curve in Fig. 4b).

In order to measure the time response of the active MZI device, the laser was tuned to $1,630\ \text{nm}$ wavelength, where the application of $2\ \text{mW}$ power to the signal arm results in almost π phase shift, as is evident from corresponding spectra in Fig. 4a (black and red curves). Transmission through the MZI is switched from the opaque state (fringe minimum in the black spectrum in Fig. 4a) to almost complete transmission (fringe maximum in the red spectrum in Fig. 4a) within only $100\ \text{ns}$, as is seen from Fig. 4c. This time response is the fastest yet reported for thermo-optic modulators^{22–25}, and is a direct result of the deep scaling of the device size. In our device, the heat generation and dissipation is localized to a very small volume in the silicon photonic crystal membrane (estimated as only $400\ \mu\text{m}^3$).

METHODS

Design and fabrication of the unbalanced passive photonic crystal MZI.

Photonic crystal (PhC) structures were fabricated on $10\ \Omega\ \text{cm}$ p-type silicon-on-insulator $200\ \text{mm}$ wafers from SOITEC with a $2\ \mu\text{m}$ buried oxide layer (BOX) and a thin silicon layer (thickness $d = 223\ \text{nm}$) on a standard CMOS fabrication line as described elsewhere^{11,12}. PhCs with a triangular lattice of holes with a period $a = 437\ \text{nm}$ were defined by etching holes with radius $R \approx 0.25a$ ($109\ \text{nm}$) through a silicon layer. Note that to create an unbalanced MZI, the PhC waveguide in the reference arm and signal arm have holes of radius $0.251a$ and $0.263a$, respectively. PhC waveguides were formed by omitting one row of holes in the lattice along the Γ –K direction, as shown in Fig. 1a. After etching the silicon, the BOX was selectively etched away in the PhC waveguide region leaving a suspended silicon membrane perforated with holes (Fig. 1b). PhC waveguides with these parameters are known to possess a very broad bandwidth, with the low loss single-mode propagation extending from $1,520\ \text{nm}$ to $1,620\ \text{nm}$ with losses measured as small as $8 \pm 2\ \text{dB cm}^{-1}$ (ref. 12). Access strip waveguides with $460\ \text{nm}$ width are butt-coupled to the PhC waveguides through a lateral taper with the final width of $757\ \text{nm}$ corresponding to $\sqrt{3}a$ (ref. 11), which closely matches the geometrical spread of the modes. In order to reduce the coupling losses arising from strong mode impedance mismatch at the coupling interface¹⁹, the PhC lattice was terminated at the position of the cut going through the centre of the holes nearest to the waveguide, as shown in Fig. 1a. Utilization of efficient ultra-compact 15° angle Y-splitters and 90° bends with radius $R = 5\ \mu\text{m}$ (ref. 28) (see Fig. 1b, c) allows not only the footprint of the MZI device to be shrunk to $0.04\ \text{mm}^2$, but also allows very low insertion losses through the MZI; below $1\ \text{dB}$ over a wide bandwidth, as evident from Fig. 2a. In principle, the device footprint can be easily scaled down to $0.01\ \text{mm}^2$ and is currently limited by a necessity to place large electric contacts inside the MZI for easy electric probing. In order to couple light from a fibre to the photonic chip, the access strip waveguides were

terminated at the input and output with polymer fibre couplers based on an inverted taper design, as reported in ref. 11.

Design and fabrication of an electrically tunable PhC MZI. Active MZI devices were created by adding lateral electrical contacts to the PhC waveguide. Ohmic contacts were formed by p-doping the outer edge of the PhC region and contact pad regions outside the MZI with boron ions to $10^{20}\ \text{cm}^{-3}$ concentration, as shown in Fig. 3b. Metal contacts were formed in the same region by nickel deposition and subsequent silicidation. In order to decrease the device impedance, the central core of the PhC waveguide was also doped, to a concentration of only $10^{17}\ \text{cm}^{-3}$. This provides impedances in the $10\ \text{k}\Omega$ range while increasing the insertion loss by only $2\ \text{dB}$, as measured from corresponding transmission spectra in Fig. 4a. Fabrication of the finished MZI device requires six levels of lithography. A detailed description and schematic of the fabrication sequence is presented in Supplementary Information.

Measurements of transmission spectra. TE-polarized (electric field in the silicon slab plane) transmission spectra were measured with two complementary set-ups. The first utilizes a broad band light source covering the $1,200$ – $1,700\ \text{nm}$ wavelength region (Agilent 83437B), and spectra were recorded with an optical spectrum analyser (Agilent 86140B) with $60\ \text{pm}$ resolution (blue curve in Fig. 2a and all spectra in Fig. 4a). The second set-up uses solid state lasers (New Focus, Velocity 6328H, 6330), tunable in $0.12\ \text{pm}$ steps in the $1,500$ – $1,632\ \text{nm}$ wavelength range (red spectra in Fig. 2a). The narrow linewidth of the laser ($100\ \text{MHz}$) corresponds to a coherence length that exceeds all characteristic optical lengths in the measurement system, and allows accurate resolution of interference fringes with a free spectral range below $0.01\ \text{nm}$. The light sources are coupled to a polarization maintaining (PM) fibre and are passed to the polarization controller. Coupling to/from the photonic chip is provided via micro-lensed and tapered PM fibres aligned with the input/output on-chip polymer fibre couplers using xyz piezo-translational stages^{11,12}. Efficient coupling ($<0.5\ \text{dB}$ loss per port¹¹) allows the mean amplitude of the Fabry–Perot noise to be minimized to only $1.5\ \text{dB}$. A rejection ratio of over $30\ \text{dB}$ between TE and TM polarizations is achieved. The transmission spectra of the PhC waveguides and the MZI devices were normalized to the transmission spectra of corresponding single-mode reference strip waveguides¹² located in the same writing field near the PhC structures¹¹.

Numerical simulations. To obtain the dispersion of the group index, the photonic band structure was calculated with a three-dimensional full-vector plane wave expansion method using the MIT Photonics Band software code²⁰. The dielectric permittivity of the Si slab was taken as 12.1 . Values for the hole radii, lattice constant and slab thickness were taken from the scanning electron microscope (SEM) measurements. The grid resolution (number of vectors in the unit cell of the PhC) of the plane wave calculations was set to $16 \times 16 \times 16$, which gives an error in the eigenvalue convergence below $\sim 2\%$ with reasonable calculation time.

Received 5 April; accepted 6 September 2005.

- Brillouin, L. *Wave Propagation and Group Velocity* (Academic, New York, 1960).
- Hau, L. V. et al. Light speed reduction to 17 meters per second in an ultracold atomic gas. *Nature* **397**, 594–598 (1999).
- Liu, C. et al. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
- Bajcsy, M. et al. Stationary pulses of light in an atomic medium. *Nature* **426**, 638–641 (2003).
- Lukin, M. D. & Imamoglu, A. Controlling photons using electromagnetically induced transparency. *Nature* **413**, 273–276 (2001).
- Ku, P. C. et al. Slow light in semiconductor quantum wells. *Opt. Lett.* **29**, 2291–2293 (2004).
- Turukhin, A. V. et al. Observation of ultraslow and stored light pulses in a solid. *Phys. Rev. Lett.* **88**, 023602 (2002).
- Bigelow, M. S. et al. Superluminal and slow light propagation in a room-temperature solid. *Science* **301**, 200–202 (2003).
- Mok, J. T. & Eggleton, B. J. Expect more delays. *Nature* **433**, 811–812 (2005).
- Khurgin, J. B. Optical buffers based on slow light in electromagnetically induced transparent media and coupled resonator structures: comparative analysis. *J. Opt. Soc. Am. B* **22**, 1062–1074 (2005).
- McNab, S. J., Moll, N. & Vlasov, Y. A. Ultra-low loss photonic integrated circuit with membrane-type photonic crystal waveguides. *Opt. Express* **11**, 2927–2939 (2003).
- Dulkeith, E., McNab, S. J. & Vlasov, Y. A. Mapping the optical properties of slab-type two-dimensional photonic crystal waveguides. *Phys. Rev. B* **72**, 115102 (2005).
- Krauss, T. F., De La Rue, R. & Brand, S. Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
- Notomi, M. et al. Extremely large group-velocity dispersion of line-defect waveguides in photonic crystal slabs. *Phys. Rev. Lett.* **87**, 253902 (2001).

15. Vlasov, Y. A. *et al.* Femtosecond measurements of the time of flight of photons in a three-dimensional photonic crystal. *Phys. Rev. E* **60**, 1030–1035 (1999).
16. Gersen, H. *et al.* Real-space observation of ultraslow light in photonic crystal waveguides. *Phys. Rev. Lett.* **94**, 073903 (2005).
17. Astratov, V. N. *et al.* Heavy photon dispersions in photonic crystal waveguides. *Appl. Phys. Lett.* **77**, 178–180 (2000).
18. Altug, H. & Vuckovic, J. Experimental demonstration of the slow group velocity of light in two-dimensional coupled photonic crystal microcavity arrays. *Appl. Phys. Lett.* **86**, 111102 (2004).
19. Vlasov, Y. A. & McNab, S. J. Coupling into the slow light mode in slab-type photonic crystal waveguides. *Opt. Lett.* (in the press); preprint at (<http://arxiv.org/abs/physics/0504102>) (2005).
20. Johnson, S. G. & Joannopoulos, J. D. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express* **8**, 173–190 (2001).
21. Bendickson, J. M., Dowling, J. P. & Scalora, M. Analytic expressions for the electromagnetic mode density in finite, one-dimensional, photonic band-gap structures. *Phys. Rev. E* **53**, 4107–4121 (1996).
22. Cocorullo, G. & Rendina, I. Thermo-optical modulation at 1.5 μm in silicon etalon. *Electron. Lett.* **28**, 83–85 (1992).
23. Geis, M. W., Spector, S. J., Williamson, R. C. & Lyszczarz, T. M. Submicrosecond submilliwatt silicon-on-insulator thermooptic switch. *IEEE Photon. Technol. Lett.* **16**, 2514–2516 (2004).
24. Espinola, R. L., Tsai, M.-C., Yardley, J. T. & Osgood, R. M. Fast and low-power thermooptic switch on thin silicon-on-insulator. *IEEE Photon. Technol. Lett.* **15**, 1366–1368 (2003).
25. Camargo, E. A., Chong, H. M. H. & De La Rue, R. M. 2D Photonic crystal thermo-optic switch based on AlGaAs/GaAs epitaxial structure. *Opt. Express* **12**, 588–592 (2004).
26. McNab, S. J., Hamann, H. F., Vlasov, Y. A. Lateral electrical contacts for photonic crystal based integrated opto-electronic devices. Pending US patent US20050084195A1 (15 October 2003).
27. McNab, S. J., Hamann, H. F., O'Boyle, M. & Vlasov, Y. A. Method and apparatus for thermo-optic modulation of optical signals. Pending US patent, US20050084213A1 (12 January 2004).
28. Vlasov, Y. A. & McNab, S. J. Losses in single-mode silicon-on-insulator strip waveguides and bends. *Opt. Express* **12**, 1622–1631 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was partially supported by the DARPA DSO Slow Light programme.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.A.V. (yvlasov@us.ibm.com).

LETTERS

Simulating micrometre-scale crystal growth from solution

Stefano Piana¹, Manijeh Reyhani¹ & Julian D. Gale¹

Understanding crystal growth is essential for controlling the crystallization used in industrial separation and purification processes. Because solids interact through their surfaces, crystal shape can influence both chemical and physical properties¹. The thermodynamic morphology can readily be predicted², but most particle shapes are actually controlled by the kinetics of the atomic growth processes through which assembly occurs³. Here we study the urea–solvent interface at the nanometre scale and report kinetic Monte Carlo simulations of the micrometre-scale three-dimensional growth of urea crystals. These simulations accurately reproduce experimentally observed crystal growth. Unlike previous models of crystal growth^{4–6}, no assumption is made that the morphology can be constructed from the results for independently growing surfaces or from an *a priori* specification of surface defect concentration. This approach offers insights into the role of the solvent, the degree of supersaturation, and the contribution that extended defects (such as screw dislocations) make to crystal growth. It also connects observations made at the nanometre scale, through *in situ* atomic force microscopy, with those made at the macroscopic level. If extended to include additives, the technique could lead to the computer-aided design of crystals.

The diversity of crystal morphologies that can be found for a single material is testimony to the fact that the macroscopic shape is highly sensitive to the growth conditions, because kinetic control is usually dominant⁷. For example, Fig. 1a and b illustrates two distinct morphologies that are exhibited by the molecular crystalline material urea—(NH₂)₂CO—which one is obtained depends on whether the solvent used is water or methanol. There are even variations between individual particles, affected by their age and by when they successfully nucleated relative to other crystallites. The challenge is to be able to predict such behaviour and to reconcile it with atomic detail, such as is becoming available from *in situ* scanning probe microscopy. Here we will demonstrate that computer simulation can provide a means of bridging this gap and make it feasible to explore the crystal growth process with almost no prior assumptions.

It was previously⁶ shown that the broad features of the morphology of urea in water could be determined by information obtained from the molecular dynamics simulation of the aqueous interface for both the (001) and (110) surfaces. It was assumed that the system was growing at low supersaturation and therefore that screw dislocations would be the dominant growth site for all faces. Hence, the relative rates of growth are determined on the basis of the thermodynamics of incorporating molecules at kink sites. Now, with the advance of computer power, it is possible⁸ to determine the rates directly for all the steps of growth of the urea surface. To achieve this involves classifying the individual urea molecules as being either crystalline or in solution. Crystalline sites are then subdivided according to the local coordination environment of the molecule, based on the number of neighbouring molecules of a given type, as

shown in Fig. 2. By literally counting the number of transitions between different sites during several nanoseconds of simulation we can obtain the rate for each step. Note that rates are determined for both the dissolution step, in which a molecule moves from a given

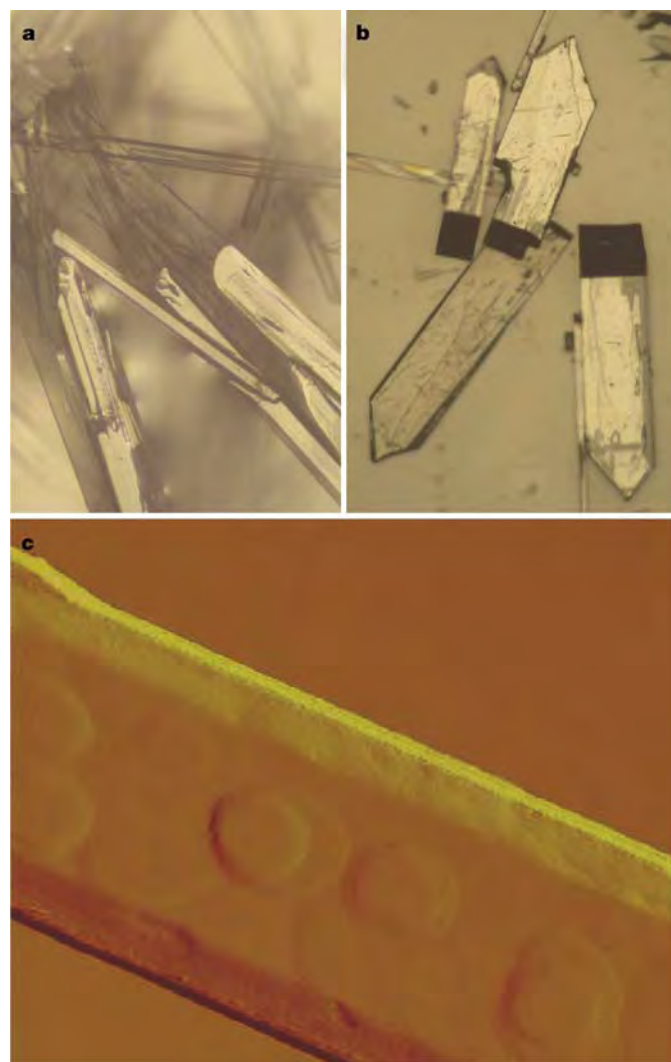


Figure 1 | Optical microscope and scanning probe microscope pictures of urea crystals. **a**, Needle-like morphology characteristic of urea crystals growing from water solution. **b**, Polar morphology of urea crystals growing from a methanol solution. **c**, Scanning probe microscope image of the [110] face of a methanol-grown urea crystal, displaying a morphology typical of a birth-and-spread growth mechanism.

¹Nanochemistry Research Institute, Department of Applied Chemistry, Curtin University of Technology, GPO Box U1987, Perth 6845, Western Australia.

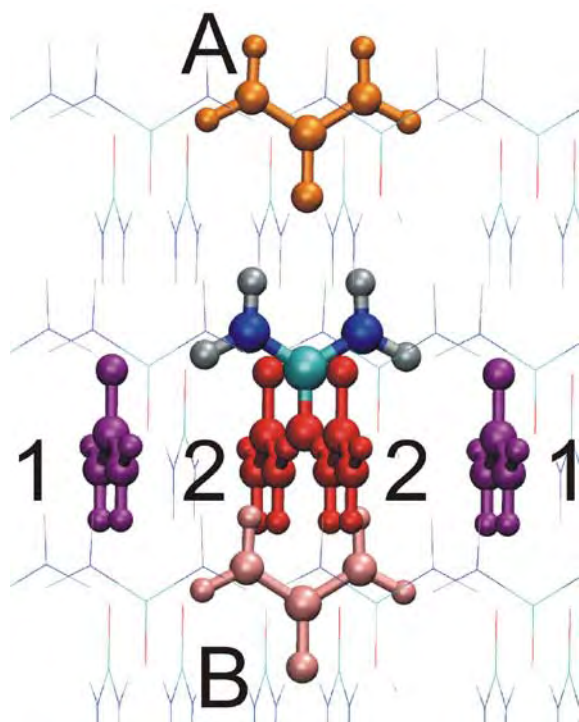


Figure 2 | Characterization of the local environment of a urea molecule. The four symmetry non-equivalent neighbouring molecules to a given urea molecule are illustrated, where the molecules labelled 1 and 2 lie in the same crystallographic *a–b* plane, whereas the molecules A and B lie above and below, respectively, along the *c* axis.

crystalline environment to solution, and the crystallization step, which is the reverse process. Examination of the influence of the classification criteria and the sampling methods used indicates that the results are not particularly sensitive to these factors, provided the simulation duration is sufficient.

Here we have simulated the solvent–urea interface for both water and methanol, as well as for four different surfaces, namely the (001), (110), (111) and $(-1-1-1)$ cuts. We note that the polar (111) and $(-1-1-1)$ surfaces are non-equivalent. Snapshots from the molecular dynamics trajectories are shown in Fig. 3. In the present methodology, it is not necessary to simulate all possible surfaces to obtain reliable results, as would be the case when determining the thermodynamic morphology according to a Wulff construction⁹. It is sufficient to determine rates for all possible transitions between unique sites that might occur on any surface. In the case of urea, the higher index surfaces can be considered to be faceted versions of the surfaces studied here, so we already have a complete set of surface sites within a nearest-neighbour and next-nearest-neighbour model.

So far, only rates for molecular processes at the nanoscale have been determined because they are extracted from simulations where the system dimensions are of the order of nanometres and the amount of real time sampled is up to 100 ns. However, the smallest observable crystallites in an optical microscope are on the micrometre scale—that is, three orders of magnitude larger—and the timescale for observable growth of a urea crystal is of the order of milliseconds. With current computers it is impractical to simulate crystal growth directly at a macroscopic level. However, by using the rates for individual steps as the probabilities for transitions in a kinetic Monte Carlo¹⁰ simulation it is feasible to make the connection between nanoscale simulation, scanning probe microscopy and the observed morphology of crystals from optical microscopy.

In the kinetic Monte Carlo approach used, each molecule is represented by a point on a grid with the arrangement of the lattice

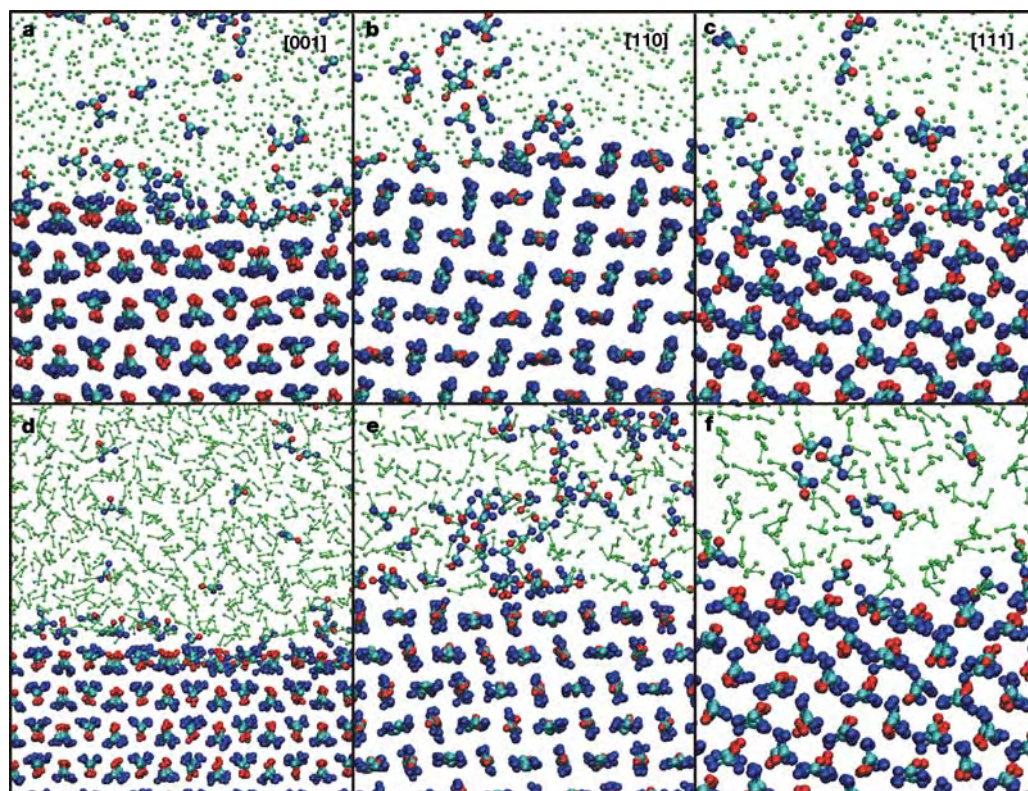


Figure 3 | Snapshots from the molecular dynamics simulations of the urea–solvent interface. **a–c**, For water, the frames show the [001] (**a**), [110] (**b**), and [111]– $(-1-1-1)$ (**c**) surfaces. **d–f**, As for **a–c** but for methanol.

sites of the urea crystal structure. Each site then carries the information of whether it is occupied by a urea molecule, or vacant, and whether the molecular dipole is pointing up or down. The orientation of the dipole is equivalent to defining the plane in which the urea molecule is situated. Sequentially, each site is examined and a pathway is chosen in proportion to the transition probability over a 50-ps interval. Here the time interval is set equal to that used to sample the transitions within the atomistic molecular dynamics. Once all sites have been considered, the clock is advanced by 50 ps and the process repeated. The present method differs from conventional approaches to kinetic Monte Carlo simulations in one important respect: the rates are obtained directly by simulation, whereas normally activation energies are calculated and then rates are estimated based on an approximation of the prefactor¹¹. In the context of crystal growth, the present work also deviates from previous studies¹² using kinetic Monte Carlo simulations by considering the full three-dimensional evolution of a crystal, rather than two-dimensional growth of a particular surface.

To initialize the simulations, a growth nucleus is created that is larger than the critical size. The shape can be chosen arbitrarily because during the initial stages the nucleus undergoes dissolution and growth until the stable morphology is achieved. Thereafter crystal growth begins, depending on the degree of supersaturation, C^* , of the urea solution surrounding the nucleus. Here C^* is defined as $([\text{urea}] - [\text{urea}]_s)/[\text{urea}]_s$, where $[\text{urea}]_s$ is the saturated solution concentration. Interestingly, for small crystal nuclei, ranging in size from 0.12 to 0.01 μm , it was found that $[\text{urea}]_s$ becomes strongly

size-dependent and for the small nuclei (~ 10 nm radius) can be up to 20% higher than the bulk limit.

To control the degree of supersaturation, the fraction of sites that are occupied by a urea molecule in solution can be manipulated. Two sets of conditions can be created—one in which the number of urea molecules is finite, and so C^* decreases with time until growth is halted, and one in which C^* is held constant by replenishing the reservoir of urea molecules as the number is depleted by crystal growth.

Simulations of urea crystals growing from methanol and water solutions at a supersaturation of $C^* = 3 \times 10^{-2}$ were performed. Figure 4a and b shows a comparison of the crystal morphologies obtained from these simulations with the morphology of urea crystals observed with an *in situ* optical microscope. The kinetic Monte Carlo simulation in water indicates that urea crystals grow from aqueous solution as long needles, of aspect ratio greater than 1,000, with large [110] faces and a slightly faceted [001] face. Growth on the [110] faces is not observed at this supersaturation, in agreement with the experimental observation that at a supersaturation of $\sim 10^{-3}$, the growth on the [110] face is more than three orders of magnitude slower than that on the [001] face¹³. In contrast, growth on the [001] face proceeds via a birth-and-spread mechanism where nucleation is not rate-limiting (rough growth).

The morphology emerging from the simulations of growth from a methanol solution is remarkably different from the simulations in water solution. The aspect ratio of methanol-grown crystals is 20–100 times smaller and the [001] faces are unstable, being completely replaced by polar [111] faces (Fig. 4b). All these findings are completely consistent with the optical microscope and scanning probe microscopy images of urea crystals growing from methanol solution (Fig. 1). We find nucleation to be rate-limiting for growth on the [110] face at low to moderate supersaturation ($C^* < 10^{-1}$). This growth mechanism has been confirmed by scanning probe microscopy data (Fig. 1c), in which the surface morphology typical of a birth-and-spread growth mechanism can be clearly identified. We therefore predict the aspect ratio to be size-dependent under these conditions, with values in the range of 10–40.

From the simulations, we can observe exactly why urea exhibits the polar morphology in which the two ends are capped by points twisted by 90 degrees. This is due to the rate of growth of steps on the [110] face from methanol being different in the $+c$ and $-c$ directions (Fig. 4b). On the symmetry-equivalent [110] and $[-1 - 10]$ faces, the rate of molecular deposition on steps in the $+c$ direction is twice that in the $-c$ one, while for the $[1 - 10]$ and $[-110]$ faces, the situation is reversed. This asymmetry in the growth rate is exhibited by islands growing on the surface, because the initial nucleation site remains close to one edge of the island, rather than being located at the centre.

It has been proposed^{6,13} that, at low supersaturation, the presence of screw dislocations is important for the growth of the [001] face. To investigate this issue, we performed two-dimensional kinetic Monte Carlo simulations of a $0.1 \times 0.1 \mu\text{m}$ [001] face, at a supersaturation of 1×10^{-3} , with and without a screw dislocation present. The presence of a screw dislocation is found to accelerate the rate by only 10%. However, on the nucleation-limited [110] face, the introduction of a screw dislocation indeed produces a fivefold enhancement of the growth rate, up to a supersaturation of 3×10^{-2} .

Our results show that it is possible to explore the influence of supersaturation, concentration gradients and diffusion control within the solvent, to name only a few of the many factors that influence crystal growth. Critical nuclei sizes can also readily be determined as a function of conditions. With further atomistic simulations, we may next examine the interplay between impurities, supersaturation and crystal growth¹⁴. Although it may prove more complex and computationally demanding, this approach is applicable to crystals in general, including species with conformational freedom. Three-dimensional kinetic Monte Carlo simulation, based

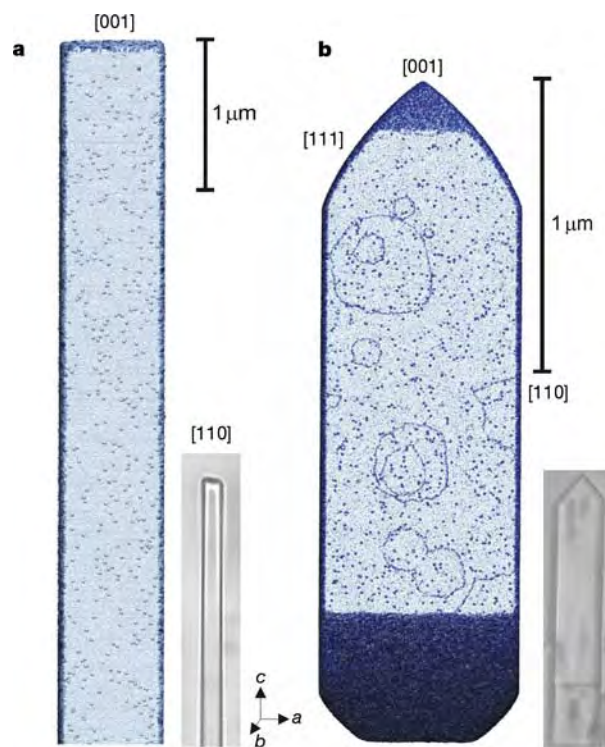


Figure 4 | Comparison of the experimental and theoretical micrometre-scale crystal morphologies. The frames show the results of the kinetic Monte Carlo simulation of a urea crystal growing from solution and the comparable *in situ* optical microscope image for both water (a) and methanol (b) as the solvent. In the simulated images the shade of a site depends on the local coordination number. Dark, medium and light blue sites are those with three, four and five neighbours, respectively. The birth-and-spread growth patterns observed in the kinetic Monte Carlo simulation of growth from a methanol solution can be compared with the scanning probe microscopy data in Fig. 1c.

on rates obtained directly from nanoscale simulation, thus provides a new technique for the predictive design of crystal growth experiments.

METHODS

More details of the methodology are presented in the Supplementary Information.

Atomistic molecular dynamics simulations. Molecular dynamics simulations were performed for the [001], [110] and [111]/[−1−1−1] crystal faces of urea in contact with water (w) and methanol (m) solutions. The initial coordinates were generated with the program GDIS¹⁵ from the unit cell determined by X-ray diffraction¹⁶. The surface area was 8 × 8, 6 × 6 and 8 × 8 unit cells for the [001], [110] and [111] surfaces, respectively. The depths of the urea slabs were 6, 8 and 5 unit cells, respectively. The two-dimensional cells were converted into three-dimensional cells having the *c* axis perpendicular to the surface and magnitude 25 Å larger than the unit cell. The gap between the two surfaces was filled with solvent molecules using the genbox¹⁷ program. The final systems consisted of 768, 576 and 320 urea molecules and 1,295, 931 and 631 solvent molecules. All the simulations were run with the program GROMACS¹⁷, using force fields for the urea–urea and urea–solvent interactions previously parameterized^{18,19}. The particle mesh Ewald²⁰ method was used for the long-range electrostatics with a short-range cut-off of 0.9 nm. The time step for the molecular dynamics simulation was 2.0 fs. Further details on the methodology can be found elsewhere⁸. Solvent molecule positions were first relaxed by geometry optimization and then the density of the system was equilibrated by performing 300 ps of molecular dynamics simulation at 300 K with a variable cell along the *c* axis and the urea molecules fixed. The whole system was equilibrated by performing 200 ps of molecular dynamics simulation at 150 K followed by 800 ps at 300 K. Finally, six simulations of 12, 60 and 12 ns in duration were performed for the [001], [110] and [111]/[−1−1−1] surfaces, respectively, with anisotropic pressure coupling.

Kinetic Monte Carlo simulations. Given the symmetry of the urea crystal, each molecule has four non-equivalent nearest-neighbour sites, illustrated in Fig. 2. When allowing for the possibility of solvent molecules occupying one or more of these positions, this leads to a total of 34 different surface sites types (ten types of kink, eight types of step, four types of terrace sites and 12 types of molecule with one or two urea neighbours only). For many surface sites the second-nearest neighbours have a very limited influence on the reaction rates (less than the statistical error) and so a nearest-neighbour scheme is normally used. However, this is not true for both the step and kink sites, where second-nearest-neighbour information was taken into account in determining reaction rates for these sites.

The calibration of the urea concentration for a saturated solution, [urea]_s, was obtained by performing a kinetic Monte Carlo simulation of a screw dislocation on a two-dimensional periodic 0.6 × 0.6 μm [110] surface at variable concentration. In this simulation a fixed number of molecules are present in the solution and the concentration is depleted as the surface grows around the screw dislocation. The saturated solution concentration was defined as the concentration where the surface neither grows nor dissolves, and was found to be 5 M and for water and 4 M for methanol, respectively. These values can be compared with the experimental values of 11 M (ref. 21) and 4 M (ref. 22), and the previously published⁸ saturated solution concentration of 7 M for water, obtained directly by molecular dynamics. The calculated rates are in qualitative agreement with the experiment and appear to capture the larger solubility of urea in water with respect to methanol. Although the simulations underestimate the dissolution of urea in water, this represents an error of only 2 kJ mol^{−1} in the free energy—an amount beyond the accuracy of most theoretical methods at present. Similarly, the discrepancy between the saturated solution concentrations for water in the kinetic Monte Carlo and the explicit simulations represents a very small error in the thermodynamic mapping.

Scanning probe microscopy. Urea crystals were imaged with an atomic force microscope (deflection plot). Crystals were grown from a saturated methanol solution by solvent evaporation and deposited on a mica surface. A Digital Instruments atomic force microscope, operated in contact mode, was used to take images of the crystals.

Received 28 June; accepted 4 August 2005.

- Winn, D. & Doherty, M. F. Modeling crystal shapes of organic materials grown from solution. *Am. Inst. Chem. Eng. J.* **46**, 1348–1367 (2000).
- Gibbs, J. W. *Collected Works* (eds Longley, W. R. & van Name, R. G.) (Longman, New York, 1928).
- Rodriguez-Hornedo, N. & Murphy, D. Significance of controlling crystallization mechanisms and kinetics in pharmaceutical systems. *J. Pharm. Sci.* **88**, 651–660 (1999).
- Hartman, P. & Perdok, W. G. On the relations between structure and morphology of crystals. *Acta Crystallogr.* **8**, 48–52 (1955).
- Pina, C. M., Becker, U., Risthous, P., Bosbach, D. & Putnis, A. Molecular-scale mechanisms of crystal growth of barite. *Nature* **395**, 483–486 (1998).
- Liu, X. Y., Boek, E. S., Briels, W. J. & Bennema, P. Prediction of crystal growth morphology based on structural analysis of the solid-fluid interface. *Nature* **374**, 342–345 (1995).
- Davey, R. J., Mullin, J. W. & Whiting, M. J. L. Habit modification of succinic acid crystals grown from different solvents. *J. Cryst. Growth* **58**, 304–312 (1982).
- Piana, S. & Gale, J. D. Understanding the barriers to crystal growth: Dynamical simulation of the dissolution and growth of urea from aqueous solution. *J. Am. Chem. Soc.* **127**, 1975–1982 (2005).
- Wulff, G. Zur frage der geschwindigkeit des wachstums und der auflösung der krystallflächen. *Z. Krist.* **34**, 449–530 (1901).
- Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* **22**, 403–434 (1976).
- Jönsson, H. Theoretical studies of atomic scale processes relevant to crystal growth. *Annu. Rev. Phys. Chem.* **51**, 623–653 (2000).
- Boerrigter, S. X. M. et al. MONTY: Monte Carlo crystal growth on any crystal structure in any crystallographic orientation; application to fats. *J. Phys. Chem. A* **108**, 5894–5902 (2004).
- Davey, R., Fila, W. & Garside, J. The influence of biuret on the growth kinetics of urea crystals from aqueous solutions. *J. Cryst. Growth* **79**, 607–613 (1986).
- Land, T. A., Martin, T. L., Potapenko, S., Palmore, G. T. & de Yoreo, J. J. Recovery of surfaces from impurity poisoning during crystal growth. *Nature* **399**, 442–445 (1999).
- Fleming, S. D. & Rohl, A. L. GDIS: a visualization program for molecular and periodic systems. *Z. Krist.* **220**, 1–5 (2005).
- Swaminathan, S., Craven, B. M., Spackman, M. A. & Stewart, R. F. Theoretical and experimental studies of the charge density in urea. *Acta Crystallogr. B* **40**, 398–404 (1984).
- Lindhal, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Modeling* **7**, 306–317 (2001).
- Smith, L. J., Berendsen, H. J. C. & van Gusteren, W. F. Computer simulation of urea–water mixtures: A test of force field parameters for use in biomolecular simulation. *J. Phys. Chem. B* **108**, 1065–1071 (2004).
- van Gusteren, W. F., et al. *Biomolecular Simulation: the GROMOS96 Manual and User Guide* 1–1024 (Vdf Hochschulverlag AG an der ETH Zurich, Zurich, 1996).
- Essman, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
- Pinck, L. A. & Kelly, M. A. The solubility of urea in water. *J. Am. Chem. Soc.* **47**, 2170–2172 (1925).
- Boomadevi, S., Dhanasekaran, R. & Ramasamy, P. Investigations on nucleation and growth kinetics of urea crystals from methanol. *Cryst. Res. Technol.* **37**, 159–168 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Rohl for discussions, and to T. Dincer for assistance with the optical microscope. S.P. acknowledges financial support from an Australian Research Fellowship, while S.P. and J.D.G. both gratefully acknowledge the support of the Government of Western Australia through the Premiers Research Fellowship programme.

Author Contributions S.P. performed all the calculations and the optical microscopy, M.R. produced the atomic force microscope images, J.D.G. conceived the project and wrote the manuscript with S.P.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.G. (J.Gale@curtin.edu.au).

LETTERS

Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content

John A. Church^{1,2}, Neil J. White^{1,2} & Julie M. Arblaster^{3,4}

Ocean thermal expansion contributes significantly to sea-level variability and rise¹. However, observed decadal variability in ocean heat content^{2,3} and sea level⁴ has not been reproduced well in climate models⁵. Aerosols injected into the stratosphere during volcanic eruptions scatter incoming solar radiation, and cause a rapid cooling of the atmosphere^{6,7} and a reduction in rainfall^{6,8,9}, as well as other changes in the climate system⁷. Here we use observations of ocean heat content^{2,3} and a set of climate simulations to show that large volcanic eruptions result in rapid reductions in ocean heat content and global mean sea level. For the Mt Pinatubo eruption, we estimate a reduction in ocean heat content of about 3×10^{22} J and a global sea-level fall of about 5 mm. Over the three years following such an eruption, we estimate a decrease in evaporation of up to 0.1 mm d^{-1} , comparable to observed changes in mean land precipitation^{6,8,9}. The recovery of sea level following the Mt Pinatubo eruption in 1991 explains about half of the difference between the long-term rate of sea-level rise⁴ of 1.8 mm yr^{-1} (for 1950–2000), and the higher rate estimated for the more recent period where satellite altimeter data are available (1993–2000)^{4,10}.

Coupled climate models show better agreement with observations on both annual and decadal timescales when volcanic forcing is included^{11–13}. Volcanic eruptions also lead to changes in ocean heat content^{14,15} but, to date, there has been little focus on their impact on sea level, other than suggestions that they may be responsible for a component of observed decadal variability⁴.

In order to isolate the volcanic signal, we use a subset of the climate simulations completed with the Parallel Climate Model (PCM¹⁶). The PCM has an atmospheric resolution of about 2.8° by 2.8° with 18 levels in the vertical and an ocean resolution of $2/3^\circ$ to $1/2^\circ$ and 32 levels¹⁷. We use two three-member ensembles started from different points in the control run. The first ensemble has time-varying volcanic, solar, greenhouse gases, tropospheric (non-volcanic) sulphates and ozone ('VSGSuOz') forcing. The second has 'SGSuOz' forcing, with the volcanic component omitted. We also use the control simulation, in which the forcing is constant. The zonally averaged volcanic forcing¹² (indicated by atmospheric optical depth in the figures) used here for the period 1890–2000 (that is, starting after the large Krakatoa eruption of 1883) was derived in a consistent way based on the total amount of sulphate released and consideration of the seasonally varying stratospheric transport and decay.

We calculated the change in global mean sea level (GMSL), relative to the control run, resulting from ocean thermal expansion from 1890 to 2000 from monthly-averaged temperatures and salinities. For the SGSuOz (VSGSuOz) simulations, there was an increase in GMSL of about 43 mm (37 mm) over the 110 yr time span (Fig. 1a). There were two large volcanic eruptions before 1915, then a relatively quiescent period followed by a sequence of major eruptions starting in 1963, of which the three largest were Mt Agung (Indonesia, 1963),

El Chichon (Mexico, 1982) and Mt Pinatubo (Philippines, 1991). Following each eruption there was a fall in GMSL of several millimetres and an abrupt cooling of the ocean, typically within a year.

The volcanic GMSL (full ocean depth) and global ocean heat content (GOHC, upper 300 m only) signal was identified by subtracting results from pairs of simulations with and without volcanic forcing and forming the ensemble average of these differences (Fig. 1b, c). There is uncorrelated variability in the individual ensemble members associated with mismatches in signals from climate variability such as El Niño/Southern Oscillation (ENSO), but the volcanic signal is clear. GMSL and GOHC fall by about 5 mm

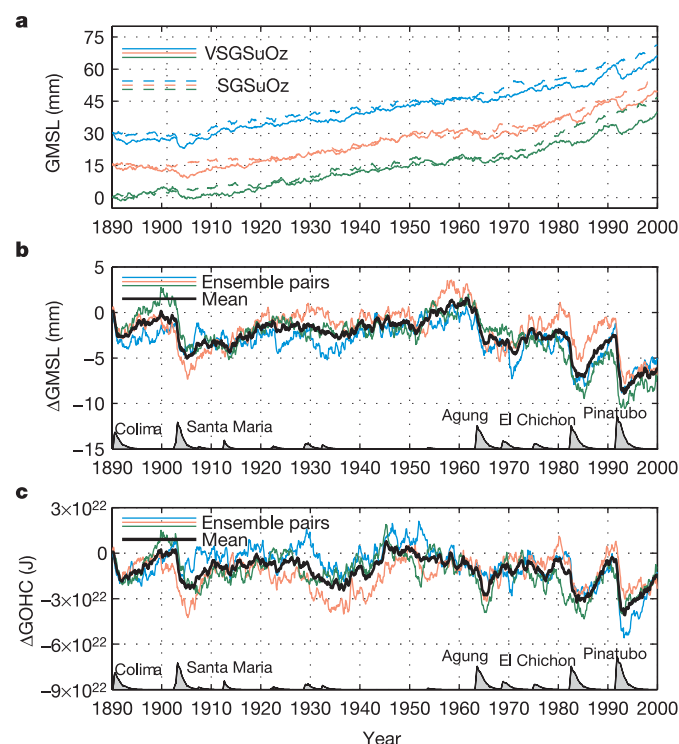


Figure 1 | Changes in global mean sea level (GMSL) and global ocean heat content (GOHC) in the PCM simulations. **a**, Three pairs of simulations of full depth GMSL with time-varying volcanic, solar, greenhouse gases, sulphates and ozone ('VSGSuOz') forcing and the corresponding simulations without the volcanic forcing ('SGSuOz') are shown respectively by the solid and dashed lines. The differences between the pairs for GMSL (ΔGMSL ; **b**) and GOHC (upper 300 m only) (ΔGOHC ; **c**) are the coloured lines and the ensemble average is the bold black line. The global average optical depth (stratospheric optical depth at $0.5 \mu\text{m}$; ref. 12) is also shown (thin black line at bottom; arbitrary scale).

¹CSIRO Marine and Atmospheric Research, GPO Box 1538, ²Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Tasmania 7001, Australia. ³National Center for Atmospheric Research, Boulder, Colorado 80307-3000, USA. ⁴Bureau of Meteorology Research Centre, Melbourne, Victoria 3001, Australia.

and 3×10^{22} J (the largest response follows the Mt Pinatubo eruption) and then rise over the following decade or more. Smaller eruptions followed the 1903 Santa Maria and the 1963 Mt Agung eruptions, and in these cases the recovery to pre-eruption levels took in excess of 15 yr. For the Mt Pinatubo eruption, the recovery was not complete at the end of the simulations in 2000. Over 1890 to 2000, the volcanic forcing leads to a reduction of about 6 mm in GMSL, most of which occurs from 1960 to 2000. This reduction offsets part of the acceleration in sea-level rise in the latter part of the twentieth century in the SGSuOz simulations.

We compare two estimates (Levitus *et al.*² and Ishii *et al.*³, updated 2005) of yearly GOHC and the associated thermal expansion component of sea level available after 1955 with the model results. We focus on the upper 300 m principally because of the better coverage of ocean data in this depth range, and because full depth and upper 300 m heat content (and sea-level) time series in the model are well correlated (for example, the correlation is 0.86 for sea level) and the variability of the pairs of time series are virtually the same magnitude. Despite this upper ocean focus, the major limitation of these ocean products is the lack of global coverage¹⁸ and reliable error estimates. The two estimates of observed GOHC and GMSL are well correlated (both above 0.8) and show a similar decrease following major volcanic eruptions. For the observations, it is not possible (as it is in the model simulations) to separate the response to the volcanic forcing from other external forcing, and the observations contain contributions from natural climate variability (for example, ENSO events) that may occur at different times to the model's internal variability. Despite the uncertainties in the volcanic forcing and the incomplete ocean database, the detrended GMSL and GOHC time series (Fig. 2) are correlated with the detrended model results, 0.59 (0.59) and 0.49 (0.44), respectively for the Levitus² (Ishii³) data sets. These post-eruption drops in sea level agree qualitatively with the observed (tide-gauge based⁴) GMSL record.

For the Mt Pinatubo eruption, for which there are better observations of stratospheric aerosols and ocean heat content, the observed GMSL and GOHC fall by about 5 mm and 3×10^{22} J, similar to the model ensemble average. This agreement supports the validity of the stratospheric aerosol loading for the Mt Pinatubo eruption¹². The slow recovery towards pre-eruption values is slightly faster in the observations than in the model. The observations are also clearly affected by the 1997/98 ENSO event.

For the 1963 Mt Agung and the 1982 El Chichon eruptions, the magnitude of the observed signal is more than twice that of the model results. For the Mt Agung eruption, the stratospheric aerosol concentration used here¹² is only 20% larger in the (ocean-dominated) Southern Hemisphere than in the Northern Hemisphere, whereas other studies suggest that it should be a factor of three to eight larger¹². To test the sensitivity of the PCM results, we use three additional climate models that use volcanic forcing. These are the twentieth-century simulations (20C3M) of the NASA Goddard Institute for Space Studies (GISS-ER) model (nine-member ensemble), and the Centre for Climate System Research, University of Tokyo, high resolution model (MIROC3.2(hires), one ensemble member) and medium resolution model¹⁹ (MIROC3.2(medres), three-member ensemble). (Documentation of these models is available from <http://www-pcmdi.llnl.gov>.) As there are no comprehensive sets of simulations available to us for these additional models that allow the specific volcanic response to be isolated, we remove the long-term sea-level rise by subtracting a quadratic (the approximate expected response to increasing greenhouse gases) from the full model depth GMSL time series for all of the models, including the PCM. The residuals (Fig. 2c) will contain the interannual volcanic signal as well as other variability. For the PCM model, the residuals are very similar to the results obtained by differencing the volcanic-non volcanic simulations. The response of all of the models is very similar (correlations between 0.7 and 0.9), and the GISS/MIROC correlations (which both use an update of the Sato *et al.*²⁰ volcanic

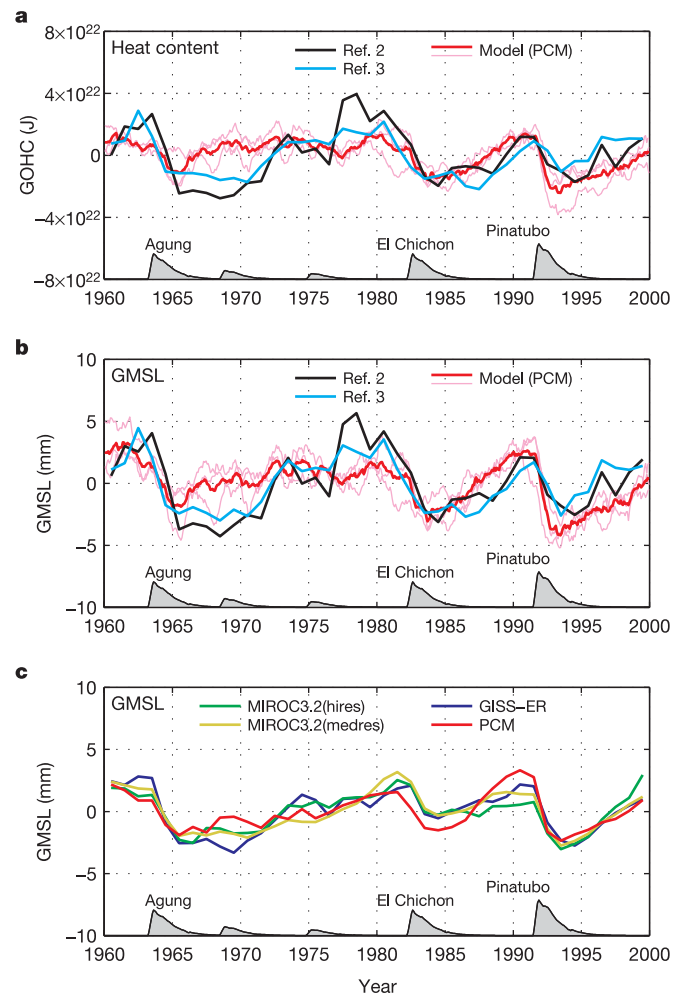


Figure 2 | Observed and modelled GOHC and GMSL for the period 1960–2000. The response to volcanic forcing, as indicated by the differences between the pairs of PCM simulations for GOHC (**a**) and the GMSL (**b**) is shown for the ensemble mean (bold line) and the three ensemble members (light lines). The observational estimates^{2,3} of GOHC and GMSL are shown by the black and blue bold lines. For **a** and **b**, all results are for the upper 300 m only and have been detrended over the period 1960–2000. **c**, The ensemble mean (full depth) GMSL for the GISS-ER, MIROC3.2(hires), MIROC3.2(medres) and the PCM models (after subtracting a quadratic) are shown.

forcing) are above 0.85. The GISS-ER model has the largest response, particularly for the Mt Agung eruption, but still smaller than the observed response. The remaining discrepancies between the observations and the model results may be due to the inadequate ocean database (particularly large gaps in the Southern Hemisphere coverage¹⁸) used for determining observed GOHC and GMSL, as well as uncertainties in the volcanic forcing and climate model sensitivity.

To better understand the processes involved, we focus on the Mt Pinatubo eruption (June 1991) because of its stronger radiative forcing and the better observations available. The primary driver of the fall in sea-surface temperature, GMSL and GOHC in the PCM is the rapid reduction in net solar flux at the ocean surface (Fig. 3a) of up to 6 W m^{-2} in late 1991. By early 1994, the net solar flux had virtually recovered to pre-eruption values, even though the optical depth from the volcanic aerosols had not yet recovered. The net shortwave forcing recovers to pre-eruption values faster than the volcanic optical depth, perhaps as a result of reduced cloud cover, as found in earlier model studies⁶. In the global average, the fall in the model sea surface temperature is about 0.4°C and almost recovers to pre-eruption values in 1995, whereas observations²¹ show a smaller

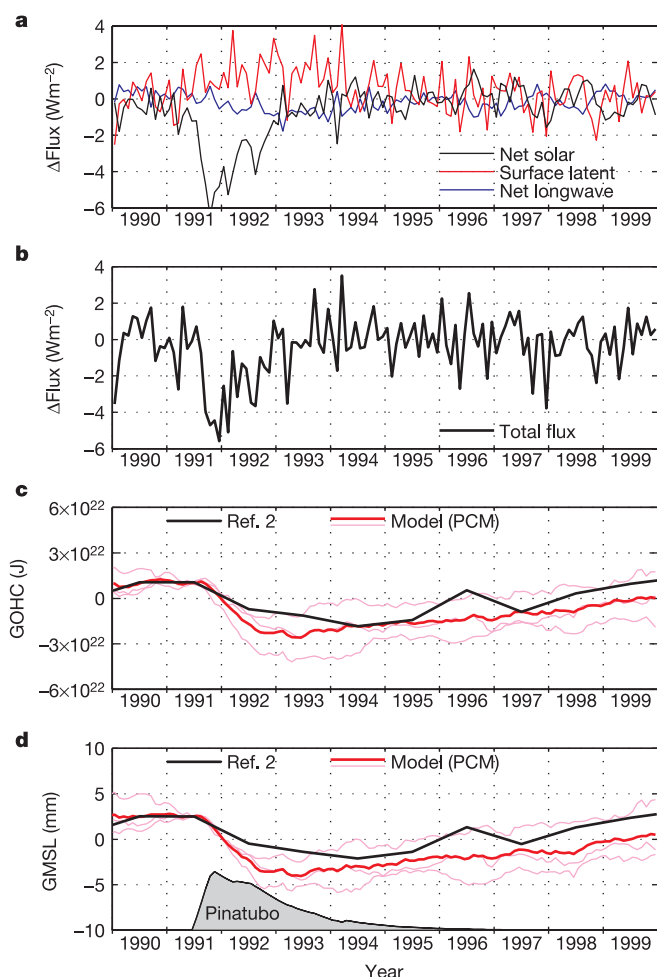


Figure 3 | Ocean heat budget anomalies associated with the Mt Pinatubo eruption. **a**, The net solar (black), latent heat (red) and net long-wave (blue) heat fluxes. **b**, The total surface heat flux. Positive (negative) fluxes are a warming (cooling) of the ocean, and the fluxes are the averages of the differences between the three ensemble pairs. **c**, The modelled and observed² GOHC. **d**, The modelled and observed² GMSL. For **c** and **d**, the three ensemble pair differences between the models with and without volcanic forcing are shown as light lines, and the ensemble average bold, and are for the upper 300 m only.

fall (about 0.3°C) and a faster recovery. In both observed and modelled sea surface temperature, the cooling occurs over a band from about 50°S to 60°N , with maximum cooling at mid-latitudes during the Southern Hemisphere summer of 1991/1992 and the Northern Hemisphere summer of 1992. The maximum cooling of both hemispheres in the summer is probably a result of shallow mixed layers at this time of year, which respond more rapidly to a given heat flux change. Sea-ice feedbacks⁶ may also be a factor.

The latent heat flux anomaly (a warming of the ocean of 2 W m^{-2} peaking in early 1993, Fig. 3a) corresponds to reduced evaporation of about 0.1 mm d^{-1} , agreeing with observed land mean precipitation reductions following volcanic eruptions^{8,9}. The maximum in the latent heat flux anomaly occurs at about the same time as the minimum of sea surface temperature and 12 months after the maximum reduction in the shortwave flux. There is also a smaller ocean cooling from the net longwave flux. The ocean cools rapidly as a result of the reduction in the total heat flux of about 5 W m^{-2} . This total flux is consistent with the observed reduction in the net forcing of the climate system between 40°N and 40°S (land and ocean)²² of 4.3 W m^{-2} . The cooling peaks in late 1991, returning to almost zero by mid-1993, then the ocean slowly warms (Fig. 3b). As a result, the

GOHC and GMSL (upper 300 m only; Fig. 3c, d) fall rapidly for 12 to 18 months following the eruption, with a slower recovery, which was not complete by 2000. The model results for the full ocean depth suggest that sea level recovers more rapidly than heat content, perhaps as a result of more rapid ocean turnover times in low latitudes where the thermal expansion coefficient is larger. In cooler, high-latitude waters with smaller thermal expansion coefficients, the heat anomalies may have been advected into the main thermocline where they would remain for decades, as found in simulations that include the larger Krakatoa (1883) and Tambora (1815) eruptions (ref. 23, and J.M. Gregory, J.A. Lowe and S.F.B. Tett, manuscript in preparation).

Quantifying the impacts of volcanic eruptions is important to understanding climate and GMSL variability, and there are important consequences for interpretation of the observational record. The rate of sea-level rise for the modern satellite altimeter era (3.2 mm yr^{-1} for 1993–2000)^{4,10} is significantly larger than the 1950–2000 rate⁴ of 1.8 mm yr^{-1} . The PCM model results indicate that the rate of sea-level rise calculated for 1993–2000 should be about 0.5 mm yr^{-1} higher than the average rate of sea-level rise over the preceding four decades, because of the recovery of sea level from the effects of the Mt Pinatubo eruption. This recovery, together with recent increases in glacier and ice sheet contributions^{24–28} (greater than 0.5 mm yr^{-1}), explains much of the difference between the 1950–2000 and 1993–2000 estimates of sea-level rise.

Received 8 June; accepted 27 September 2005.

- Church, J. A. *et al.* in *Climate Change 2001: The Scientific Basis* (ed. Houghton, J. T.) 639–694 (Cambridge Univ. Press, Cambridge, 2001).
- Levitus, S., Antonov, J. I. & Boyer, T. P. Warming of the World Ocean, 1955–2003. *Geophys. Res. Lett.* **32**, L02604 (2005) doi:10.1029/2004GL021592.
- Ishii, M., Kimono, M. & Kachi, M. Historical ocean subsurface temperature analysis with error estimates. *Mon. Weath. Rev.* **131**, 51–73 (2003).
- Church, J. A., White, N. J., Coleman, R., Lambeck, K. & Mitrovica, J. X. Estimates of the regional distribution of sea-level rise over the 1950 to 2000 period. *J. Clim.* **17**, 2609–2625 (2004).
- Gregory, J. M. *et al.* Comparison of results from several AOGCMs for global and regional sea-level change 1900–2100. *Clim. Dyn.* **18**, 225–240 (2001).
- Robock, A. & Liu, Y. The volcanic signal in Goddard Institute for Space Studies three-dimensional model simulations. *J. Clim.* **7**, 44–55 (1994).
- Robock, A. Volcanic eruptions and climate. *Rev. Geophys.* **38**, 191–219 (2000).
- Gillett, N. P., Weaver, A. J., Zwiers, F. W. & Wehner, M. F. Detection of volcanic influence on global precipitation. *Geophys. Res. Lett.* **31**, L12217, doi:10.1029/2004GL020044 (2004).
- Lambert, F., Stott, P. A., Allen, M. R. & Palmer, M. A. Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.* **31**, L10203, doi:10.1029/2004GL019545 (2004).
- Leuliette, E. W., Nerem, R. S. & Mitchum, G. T. Calibration of TOPEX/Poseidon and Jason altimeter data to construct a continuous record of mean sea level change. *Mar. Geod.* **27**, 79–94 (2004).
- Stott, P. A. *et al.* External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137 (2000).
- Ammann, C. M., Meehl, G. A. & Washington, W. M. A monthly and latitudinally varying volcanic forcing dataset in simulations of the 20th century climate. *Geophys. Res. Lett.* **30**, 16257, doi:10.1029/2003GL016875 (2003).
- Broccoli, A. J. *et al.* Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.* **108**, 4798, doi:10.1029/2003JD003812 (2003).
- Sun, S. & Hansen, J. E. Climate simulations for 1951–2050 with a coupled atmosphere-ocean model. *J. Clim.* **16**, 2807–2826 (2003).
- Hansen, J. E. *et al.* Earth's energy imbalance; confirmation and implications. *Science* **308**, 1431–1435 (2005).
- Meehl, G. A. *et al.* Combinations of natural and anthropogenic forcings in 20th century climate. *J. Clim.* **17**, 3721–3727 (2004).
- Washington, W. M. *et al.* Parallel climate model (PCM) control and transient simulations. *Clim. Dyn.* **16**, 755–774 (2000).
- Gregory, J. M., Banks, H. T., Stott, P. A., Lowe, J. A. & Palmer, M. D. Simulated and observed decadal variability in ocean heat content. *Geophys. Res. Lett.* **31**, L15312, doi:10.1029/2004GL020258 (2004).
- Nozawa, T., Nagashima, T., Shiogama, H. & Crooks, S. A. Detecting natural influence on surface air temperatures change in the early twentieth century. *Geophys. Res. Lett.* (in the press).
- Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98**, 22987–22994 (1993).

21. Rayner, N. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407, doi:10.1029/2002JD002670 (2003).
22. Minnis, P. *et al.* Radiative climate forcing by the Mount Pinatubo eruption. *Science* **259**, 1411–1415 (1993).
23. Delworth, T. L., Ramaswamy, V. & Stenchikov, G. L. The impact of aerosols on simulated ocean temperature, heat content, and sea level in the 20th century. *Geophys. Res. Lett.* (submitted).
24. Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S. & Valentine, V. B. Rapid wastage of Alaska glaciers and their contributions to rising sea level. *Science* **297**, 382–386 (2002).
25. Thomas, R. *et al.* Accelerated sea-level rise from West Antarctica. *Science* **306**, 255–258 (2004).
26. Rignot, E. *et al.* Accelerated ice discharge from the Antarctic Peninsula following the collapse of the Larsen B ice shelf. *Geophys. Res. Lett.* **31**, 118401, doi:10.1029/2004GL020697 (2004).
27. Rignot, E., Rivera, A. & Casassa, G. Contribution of the Patagonia Icefields of South America to sea level rise. *Science* **302**, 434–437 (2003).
28. Kraybill, W. *et al.* Greenland Ice Sheet: Increased coastal thinning. *Geophys. Res. Lett.* **31**, L24402, doi:10.1029/2004GL021533 (2004).

Acknowledgements This paper is a contribution to the CSIRO Climate Change Research Program. This work was supported by the Australian Government's Cooperative Research Centres Programme through the Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC). Portions of this study were supported by the Office of Biological and Environmental Research, US Department of Energy, as part of its Climate Change Prediction Program, and by the National Center for Atmospheric Research. The National Center for Atmospheric Research is sponsored by the National Science Foundation. We acknowledge the international modelling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, US Department of Energy. We thank T. Wigley for comments and insight.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.C. (John.Church@csiro.au).

LETTERS

Crustal rheology of the Himalaya and Southern Tibet inferred from magnetotelluric data

M. J. Unsworth¹, A. G. Jones², W. Wei³, G. Marquis⁴, S. G. Gokarn⁵, J. E. Spratt² & the INDEPTH-MT team*

The Cenozoic collision between the Indian and Asian continents formed the Tibetan plateau, beginning about 70 million years ago. Since this time, at least 1,400 km of convergence has been accommodated¹ by a combination of underthrusting of Indian² and Asian lithosphere, crustal shortening³, horizontal extrusion⁴ and lithospheric delamination⁵. Rocks exposed in the Himalaya show evidence of crustal melting^{1,6} and are thought to have been exhumed by rapid erosion and climatically forced crustal flow^{7,8}. Magnetotelluric data can be used to image subsurface electrical resistivity, a parameter sensitive to the presence of interconnected fluids in the host rock matrix, even at low volume fractions. Here we present magnetotelluric data from the Tibetan–Himalayan orogen from 77° E to 92° E, which show that low resistivity, interpreted as a partially molten layer, is present along at least 1,000 km of the southern margin of the Tibetan plateau. The inferred low viscosity of this layer is consistent with the development of climatically forced crustal flow in Southern Tibet.

The geology of southern Tibet clearly records the collision of India with Asia. The Indus–Tsangpo suture divides rocks of Indian and Asian origin, and the Gandese batholith to the north is a consequence of pre-collision subduction of Indian lithosphere¹. The Tethyan Himalaya to the south of the Indus–Tsangpo suture is a fold-and-thrust sequence comprising low-grade meta-sedimentary rocks deposited on the Indian continental margin of the Tethys Ocean before collision¹. In the central Tethyan Himalaya, basement windows expose gneissic domes composed of metamorphic basement rocks and small volumes of Cenozoic granitoids produced by crustal melting⁹. To the south of the Tethyan Himalaya is the Greater Himalayan Sequence composed primarily of high-grade gneisses and bounded on its upper surface by the South Tibetan detachment. The South Tibetan detachment is essentially a normal fault that places Tethyan rocks above the Greater Himalayan Sequence. The lower boundary of the Greater Himalayan Sequence is the Main Central Thrust—a largely inactive thrust whose role in orogenic convergence has been superseded by the Main Boundary Thrust and Main Frontal Thrust. The Greater Himalayan Sequence contains pervasive migmatites and is frequently intruded at the top by leucogranites that represent the product of Miocene crustal melting. Crustal melting and the extruded metamorphic slab, bounded between the Main Central Thrust and South Tibetan detachment, show that mid-crustal rocks have been exhumed in the Himalaya^{6,7}. In the northwest Himalaya the geology is similar, and the same major structural units found in southern Tibet are also present. In contrast to southern Tibet, convergence has been significantly transpressional, as expressed in the 150-km right lateral offset of the Karakorum fault¹.

Geophysical imaging in the Himalaya and Tibet has extended these surface geological studies to depth. Passive seismic data reveal a crustal thickness of up to 85 km in southern Tibet¹⁰, approximately double the global average. Seismic reflection data demonstrate that in southern Tibet this double thickness is the result of underthrusting by the Indian plate¹¹. Seismic surveys also detected bright spots that suggest a fluid phase is present at mid-crustal depths¹¹. To determine whether widespread crustal flow is occurring, information about crustal composition and rheology is required and can be inferred from complementary geophysical methods such as magnetotellurics (MT). The first MT data collected in southern Tibet detected a low-resistivity crust¹². In combination with increased heat flow¹³, it was proposed that the low resistivity was due to partial melting. INDEPTH MT data collected in 1995 and 1998 showed that the low-resistivity layer extended north from 29° N into the Lhasa block¹⁴. However, the INDEPTH survey in Southern Tibet was located within the Yadong–Gulu rift system, part of a series of Neogene rifts that accommodates east–west extension¹⁵, and it was uncertain whether the resistivity models were valid for the entire Tibetan–Himalayan orogen.

New long-period (0.05–0.0001 Hz) and broadband (300–0.001 Hz) INDEPTH MT data were collected in 2001 on the 700 and 800 lines (Fig. 1). The time series data were processed using statistically robust algorithms¹⁶. Where long-period data were available they were merged with the broadband data. The 800-line data were combined with data collected in Nepal¹⁷ to give a profile extending from Nepal to Central Tibet, and MT data from the northwest Himalaya of India¹⁸ have also been analysed. Tensor decomposition¹⁹ of the MT impedance tensors shows that the geoelectric strike is parallel to geologic strike and justifies a two-dimensional analysis (Fig. 1). The apparent resistivity decreases at frequencies below 1 Hz both north and south of the Indus–Tsangpo suture. Decreasing frequency in MT indicates an increasing depth of signal penetration. Thus this observation shows that, to first order, a low-resistivity layer is present over a significant area of southern Tibet. Representative apparent resistivity curves are shown in the online supplement.

To interpret these MT data quantitatively, it is necessary to convert frequency to depth. The electrical resistivity models shown in Fig. 2 were obtained by joint inversion of the transverse magnetic mode (electric current flow along profile) and the vertical magnetic field transfer functions with an automated algorithm²⁰. The fit of the inversion models responses to the measured MT data are shown in the Supplementary Information. A crustal low-resistivity layer is prominent in each of the models with its top at a depth of 20–25 km (pressures of 700–900 MPa) and extending south of the Indus–Tsangpo suture. The MT data can also be fitted by other resistivity

¹Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada. ²School of Cosmic Physics, Dublin Institute for Advanced Studies, 5 Merrion Square, Dublin 2, Ireland. ³Geo-detection Laboratory, Ministry of Education, China University of Geosciences, 29 Xueyuan Road, Beijing 100083, China. ⁴EOST-IPGS, EOST ULP (UMR-7516), 5 rue Rene Descartes, University of Strasbourg, Strasbourg 67084, France. ⁵Indian Institute of Geomagnetism, Colaba, Mumbai 400005, India.

*Lists of participants and affiliations appear at the end of the paper.

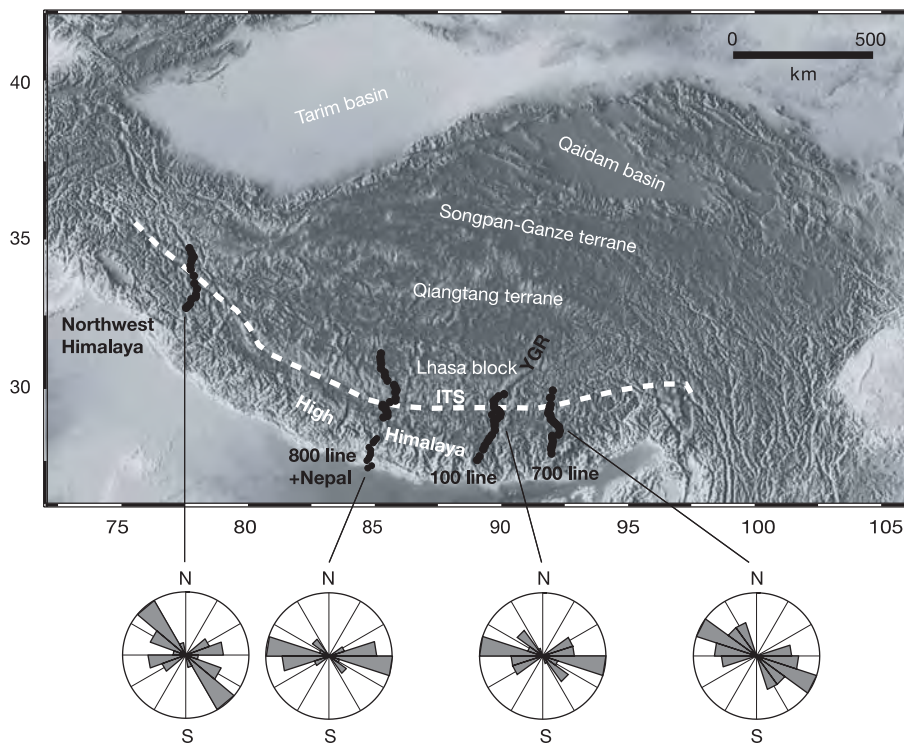


Figure 1 | Map of the Tibetan plateau showing the location of the MT profiles. The rose diagrams show geoelectric strike directions derived from the MT data using tensor decomposition. The scatter is typical of this type of data. ITS, the Indus–Tsangpo suture.

models, with the same ratio of resistivity and layer thickness. In Fig. 2 the thickest possible layer is chosen, and this corresponds to the lowest fluid fraction. The profile in the northwest Himalaya is characterized by higher mid-crustal resistivities than in southern Tibet. Differences between the models in Fig. 2 are not due solely to subsurface structure. Deep MT exploration uses natural electromagnetic signals that are variable from year to year with the solar cycle. Long-period (<1 Hz) signal levels were low in 1995 when the 100 line was recorded and high in 2001 close to the sunspot maximum when the 700 and 800 lines were recorded. The profile in the northwest

Himalaya used only broadband instruments and signal penetration was shallower. Thus the models in Fig. 2 reflect that deepest signal penetration was achieved on the 700 line. An independent analysis²¹ of the 100- and 700-line data yielded models with essentially the same primary features as in Fig. 2.

What is the origin of the low-resistivity layer? The 100 line is coincident with the INDEPTH seismic reflection profile (Fig. 3) with the Main Himalayan Thrust interpreted as the top of the under-thrusting Indian plate¹¹. An increase in resistivity is observed at the depth of the Main Himalayan Thrust, as expected for the cold Indian

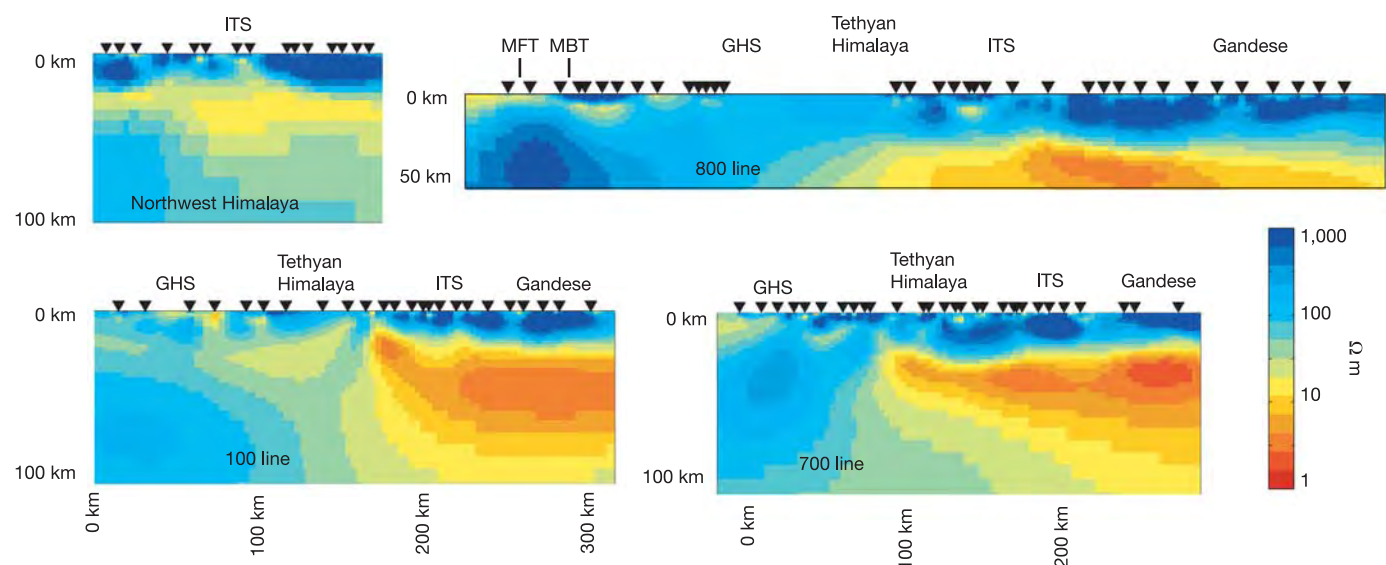


Figure 2 | Resistivity models for the four profiles derived from inversions of the MT data. The control parameters were varied to ensure that the final models were well defined. The MT data are fitted to a root-mean-square (r.m.s.) misfit in the range 1.5 and 2.5, which is statistically acceptable. Static shifts were removed from the data by allowing the inversion algorithm to

estimate the coefficients. Other approaches were used and gave consistent results. Inverted triangles denote the locations of the MT stations. MFT, Main Frontal Thrust; MBT, Main Boundary Thrust; GHS, Greater Himalayan Sequence.

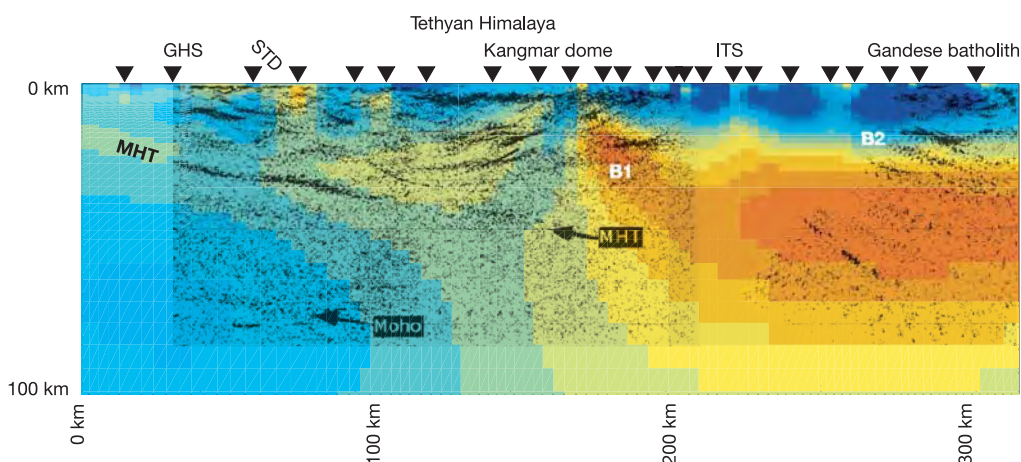


Figure 3 | Comparison of 100-line resistivity model and the INDEPTH common mid-point reflection profile. B1 and B2 are seismic bright spots that indicate zones with high fluid content. MHT, Main Himalayan Thrust; STD, Southern Tibetan detachment. Moho, Mohorovicic discontinuity.

lithosphere. The Main Himalayan Thrust reflection disappears where the low-resistivity layer begins in the Tibetan mid-crust (along-profile distance of 160 km). This can be explained if the mid-crustal layer represents a zone of high fluid content, because seismic energy will be attenuated by a fluid layer. The top of the low-resistivity layer is coincident with seismic bright spots (B1 and B2) whose seismic characteristics suggest they contain significant amounts of fluids.

The nature of these fluids is still debated, with partial melt and/or aqueous fluids as the most likely, and least contrived, explanations. Seismic reflection data suggest that the top of this layer could be aqueous fluids²², while surface wave studies suggest a broader zone characteristic of partial melting¹¹. Aqueous fluids lower the melting point of the crust and, combined with the increased heat flow in southern Tibet¹³, could cause partial melting at depths of 20–30 km (ref. 8). Thus a combination of aqueous fluids overlying a layer of partial melting gives the most consistent explanation of both the MT and seismic data^{22,23}. In addition, the geometry of the low-resistivity layer is consistent with the geometry of the zone of partial melting predicted by geodynamic models⁸. The southern edge of this zone is 50–100 km south of the Indus–Tsangpo suture and at a depth of 20–30 km. Laboratory measurements of the resistivity of hydrous granite melts gives further evidence that conditions for crustal melting occur beneath southern Tibet²⁴. Assuming that the low resistivity is primarily due to partial melting, the melt fraction was computed assuming good interconnection²⁵ and a pure melt resistivity of 0.1–0.3 Ω m (ref. 23). A bulk resistivity of 3 Ω m, typical of southern Tibet, requires a melt fraction in the range 5–14% (Fig. 4a).

Crustal flow requires a layer with a viscosity lower than the adjacent rocks and an effective viscosity below an absolute threshold that is dependent on the layer thickness. It was once believed that a melt fraction in excess of 30% was required to substantially lower the viscosity of crustal rock²⁶. However, a reexamination of laboratory data suggests that a larger, absolute reduction in viscosity occurs with a melt fraction in the range 0–7%, as the melt forms an interconnected network^{27,28,29}. When a sample of aplite was 5–7% molten, the effective viscosity was reduced by an order of magnitude²⁹. This is the amount required for strain localization and flow in geodynamic models of southern Tibet⁸.

Are these conditions encountered in Southern Tibet? The MT data require a melt fraction of 5–14%, which is sufficient to produce an order-of-magnitude reduction in viscosity and an absolute viscosity below the values necessary to account for the topography³⁰ of the Tibetan plateau (10^{16} to 10^{18} Pa s). The decrease in viscosity for granite²⁸ at these melt fractions is less, and perhaps insufficient for crustal flow to develop. It must also be noted that the extrapolation of laboratory experiments (Fig. 4b) to the low strain rates encountered in crustal deformation can be ambiguous. A similar analysis for the

northwest Himalaya yields melt fractions of 2–4% that correspond to a more modest reduction in viscosity and a less-well-developed crustal flow.

The observation of a low-viscosity mid-crustal layer has a number of geodynamic consequences. A weak layer effectively decouples the upper and lower parts of the Tibetan crust, allowing east–west

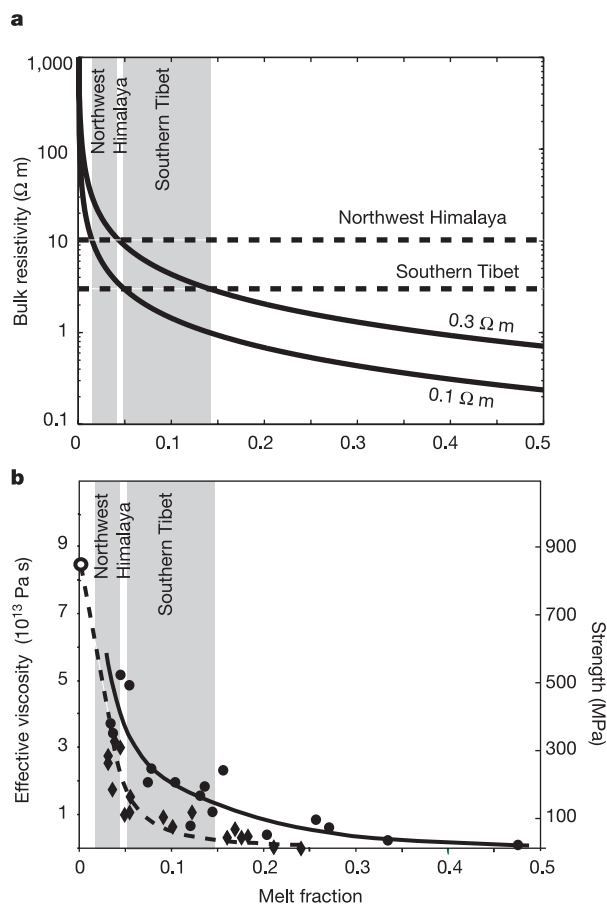


Figure 4 | Summary of laboratory measurements of the electrical resistivity and mechanical properties of a partially molten rock. **a**, Bulk electrical resistivity of partial melts as a function of melt fraction for melt resistivities of 0.1 and 0.3 Ω m; **b**, effective viscosity and strength of Westerly granite (circles) and aplite (diamonds) as a function of melt fraction²⁷. Error bars are not shown, and solid and dashed lines show best-fitting trends for Westerly granite²⁸ and aplite²⁹. The strength was computed for a strain rate of 10^{-5} s⁻¹.

extension at the surface while convergence continues at depth. The Yadong–Gulu rift¹⁵ appears to have little effect on the amount of fluid in the crust, and is probably a passive feature that has formed in response to deformation in the mid-crust. The lowest melt fractions are inferred in the northwest Himalaya, far from the extruding eastern margin of Tibet. This might be due to slower crustal motion in this area, or reflect a lower input of radiogenic heat in the underthrust rock units. There is some indication that the inferred flow may be episodic in time. Bright spot B1 probably represents an accumulation of melt, and the resistive break south of B1 may represent recently crystallized granite.

In addition to the inferred crustal flow in Southern Tibet, a larger-scale southward and eastward crustal flow may occur in northern and eastern Tibet³⁰. MT data collected in these areas have imaged zones of low crustal resistivity that may represent regions of low viscosity¹⁴.

Received 7 March; accepted 16 August 2005.

- Yin, A. & Harrison, T. M. Geologic evolution of the Himalayan–Tibetan orogen. *Annu. Rev. Earth Planet. Sci.* **28**, 211–280 (2000).
- Argand, E. in *Proceedings of the 13th International Geological Congress* 170–372 (Brussels, 1924).
- Dewey, J. F. & Burke, K. A. Tibetan, Variscan and Precambrian Basement reactivation: products of continental collision. *J. Geol.* **81**, 683 (1973).
- Tapponnier, P. et al. Oblique stepwise rise and growth of the Tibetan Plateau. *Science* **294**, 1671–1677 (2001).
- Molnar, P., England, P. & Martinod, J. Mantle dynamics, uplift of the Tibetan Plateau and the Indian monsoon. *Rev. Geophys.* **31**, 357–396 (1993).
- Le Fort, P. et al. Crustal generation of the Himalayan leucogranites. *Tectonophysics* **134**, 39–57 (1987).
- Grujic, D., Hollister, L. S. & Parrish, R. R. Himalayan metamorphic sequence as an orogenic channel: insight from Bhutan. *Earth Planet. Sci. Lett.* **198**, 177–191 (2002).
- Beaumont, C., Jamieson, R. A., Nguyen, M. H. & Lee, B. Himalayan tectonics explained by extrusion of a low-viscosity crustal channel coupled to focused surface denudation. *Nature* **414**, 738–742 (2001).
- Lee, J. et al. Evolution of the Kangmar Dome, southern Tibet: Structural, petrologic, and thermochronologic constraints. *Tectonics* **19**, 872–895 (2000).
- Owens, T. J. & Zandt, G. Implications of crustal property variations for models of Tibetan plateau evolution. *Nature* **387**, 37–43 (1997).
- Nelson, K. D. et al. Partially molten middle crust beneath southern Tibet: synthesis of project INDEPTH results. *Science* **274**, 1684–1687 (1996).
- Pham, V. N. et al. Partial melting zones in the crust in Southern Tibet from magnetotelluric results. *Nature* **319**, 310–314 (1986).
- Francheteau, J. et al. High heat-flow in southern Tibet. *Nature* **307**, 32–36 (1984).
- Wei, W. et al. Detection of widespread fluids in the Tibetan crust by magnetotelluric studies. *Science* **292**, 716–718 (2001).
- Armijo, R., Tapponnier, P., Mercier, J. L. & Han, T. L. Quaternary extension in southern Tibet: field observations and tectonic implications. *J. Geophys. Res.* **91**, 13803–13872 (1986).
- Jones, A. G., Chave, A. D., Auld, D., Bahr, K. & Egbert, G. A comparison of techniques for magnetotelluric response function estimation. *J. Geophys. Res.* **94**, 14201–14213 (1989).
- Lemmonier, C. et al. Electrical structure of the Himalaya of Central Nepal: high conductivity around the mid-crustal ramp along the Main Himalayan Thrust. *Geophys. Res. Lett.* **26**, 3261–3264 (1999).
- Gokarn, S. G., Gupta, G., Rao, C. K. & Selvaraj, C. Electrical structure across the Indus Tsangpo suture and Shyok suture zones in NW Himalaya using magnetotelluric studies. *Geophys. Res. Lett.* **29**, 1–4 (2002).
- McNeice, G. M. & Jones, A. G. Multisite, multifrequency tensor decomposition of magnetotelluric data. *Geophysics* **66**, 158–173 (2001).
- Rodi, W. & Mackie, R. L. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. *Geophysics* **66**, 174–187 (2001).
- Spratt, J., Jones, A. G., Nelson, K. D., Unsworth, M. J. & the INDEPTH MT team. Crustal structure of the India–Asia collision zone, southern Tibet, from INDEPTH MT investigations. *Phys. Earth Planet. Inter.* **150**, 227–237 (2005).
- Makovsky, Y. & Klempner, S. L. Measuring the seismic properties of Tibetan bright spots: Evidence for free aqueous fluids in the Tibetan middle crust. *J. Geophys. Res.* **104**, 10795–10825 (1999).
- Li, S. et al. Partial melt or aqueous fluids in the Tibetan crust: constraints from INDEPTH magnetotelluric data. *Geophys. J. Int.* **153**, 289–304 (2003).
- Gaillard, F., Scaillet, B. & Pichavant, M. Evidence for present-day leucogranite pluton growth in Tibet. *Geology* **32**, 801–804 (2004).
- Schilling, F. R., Partzsch, G. M., Brasse, H. & Schwarz, G. Partial melting below the magmatic arc in the central Andes deduced from geoelectromagnetic field experiments and laboratory data. *Phys. Earth Planet. Inter.* **103**, 17–31 (1997).
- Renner, J., Evans, B. & Hirth, G. On the rheologically critical melt fraction. *Earth Planet. Sci. Lett.* **181**, 585–594 (2000).
- Rosenberg, C. & Handy, M. R. Experimental deformation of partially melted granite revisited: implications for the continental crust. *J. Metamorph. Geol.* **23**, 19–28 (2005).
- Rutter, E. & Neumann, D. H. K. Experimental deformation of partially molten Westerly granite under fluid absent conditions with implications for the extraction of granitic magmas. *J. Geophys. Res.* **100**, 15697–15715 (1995).
- Van der Molen, I. & Paterson, M. S. Experimental deformation of partially molten granite. *Contrib. Mineral. Petrol.* **70**, 299–318 (1979).
- Clark, M. K. & Royden, L. H. Topographic ooze: Building the Eastern margin of Tibet by lower crustal flow. *Geology* **28**, 703–706 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The MT data were collected and analysed with support from the US National Science Foundation, the Ministry of Land and Resources of China, the Ministry of Education of China, the National Science Foundation of China, NSERC (Canada) and the Alberta Ingenuity Fund. Data in India were collected with funding from the ESS Division, Department of Science and Technology, Government of India, under the Deep Continental Studies Program. Data acquisition in Nepal was supported by CNRS-INSU and by the French–Nepalese cooperation agreement. Discussions with M. Edwards, W. Kidd and D. Nelson are acknowledged. We dedicate this paper to the memory of Doug Nelson, who inspired us all.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.U. (unsworth@phys.ualberta.ca).

The INDEPTH-MT team Paul Bedrosian¹, John Booker², Chen Leshou³, Greg Clarke⁴, Li Shenghui², Lin Chanhong³, Deng Ming³, Jin Sheng³, Kurt Solon⁵, Tan Handong³, Juanjo Ledo⁶ & Brian Roberts⁷

Affiliations for participants: ¹United States Geological Survey, ICT, MS964, Box 25046, Denver, Colorado 80225, USA. ²Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA. ³Department of Applied Geophysics, China University of Geosciences, Beijing 100083, China. ⁴Department of Earth Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada. ⁵Geological Sciences, Syracuse University, Syracuse, New York 13244, USA. ⁶Departament de Geodinàmica i Geofísica, Facultat de Geologia, Universitat de Barcelona, Martí i Franques s/n, 08028 Barcelona, Spain. ⁷Geological Survey of Canada, 615 Booth Street, Ottawa, Ontario K1A 0E9, Canada.

LETTERS

Proteorhodopsin in the ubiquitous marine bacterium SAR11

Stephen J. Giovannoni¹, Lisa Bibbs⁴, Jang-Cheon Cho^{1†}, Martha D. Stapels^{2†}, Russell Desiderio³, Kevin L. Vergin¹, Michael S. Rappé^{1†}, Samuel Laney³, Lawrence J. Wilhelm¹, H. James Tripp¹, Eric J. Mathur⁴ & Douglas F. Barofsky²

Proteorhodopsins are light-dependent proton pumps that are predicted to have an important role in the ecology of the oceans by supplying energy for microbial metabolism^{1,2}. Proteorhodopsin genes were first discovered through the cloning and sequencing of large genomic DNA fragments from seawater¹. They were later shown to be widely distributed, phylogenetically diverse, and active in the oceans^{3–7}. Proteorhodopsin genes have not been found in cultured bacteria, and on the basis of environmental sequence data, it has not yet been possible to reconstruct the genomes of uncultured bacterial strains that have proteorhodopsin genes. Although the metabolic effect of proteorhodopsins is uncertain, they are thought to function in cells for which the primary mode of metabolism is the heterotrophic assimilation of dissolved organic carbon. Here we report that SAR11 strain HTCC1062 (*Pelagibacter ubique*)⁸, the first cultivated member of the extraordinarily abundant SAR11 clade, expresses a proteorhodopsin gene when cultured in autoclaved seawater and in its natural environment, the ocean. The *Pelagibacter* proteorhodopsin functions as a light-dependent proton pump. The gene is expressed by cells grown in either diurnal light or in darkness, and there is no difference between the growth rates or cell yields of cultures grown in light or darkness.

The proteorhodopsin gene was discovered during annotation of the complete genome sequence of strain HTCC1062 (refs 9, 10). *Pelagibacter* strain HTCC1062 is a coastal, ocean surface ecotype. Previous studies have shown that the SAR11 clade occurs throughout the water column, and that it has differentiated into ecotypes that dominate in different oceanic regions and at different depths, including the deep ocean where no light penetrates¹¹.

We amplified and sequenced proteorhodopsin genes from each of ten *Pelagibacter* isolates from coastal Oregon. Amino acid substitutions were observed at two positions among the set of *Pelagibacter* proteorhodopsin genes (Supplementary Fig. S1b and Supplementary Table S1). In phylogenetic comparisons, the *Pelagibacter* proteorhodopsin genes fall within a deeply branching clade that includes genes recovered from disparate ocean water samples^{5–7} by environmental DNA cloning (Fig. 1). The *Pelagibacter* proteorhodopsin genes are most closely related to genes recovered by the shotgun cloning and sequencing of DNA from the Sargasso Sea⁷. In amino acid alignments, the *Pelagibacter* proteorhodopsin genes differ from their closest homologues in public databases at 46 out of 255 positions.

The *Pelagibacter* proteorhodopsin genes have structural features typical of rhodopsin proton pumps, including Asp 102 and Glu 113 residues, which are appropriately positioned to act as proton acceptor and donor residues in the retinylidene Schiff-base transfer

reactions that occur during the proteorhodopsin photocycle¹² (Supplementary Fig. S1b). Position 105, which is associated with spectral tuning, is occupied by a leucine residue; this is typical of shallow-water proteorhodopsins that absorb green light¹³. Proteorhodopsin genes recovered from the Sargasso Sea, although belonging to

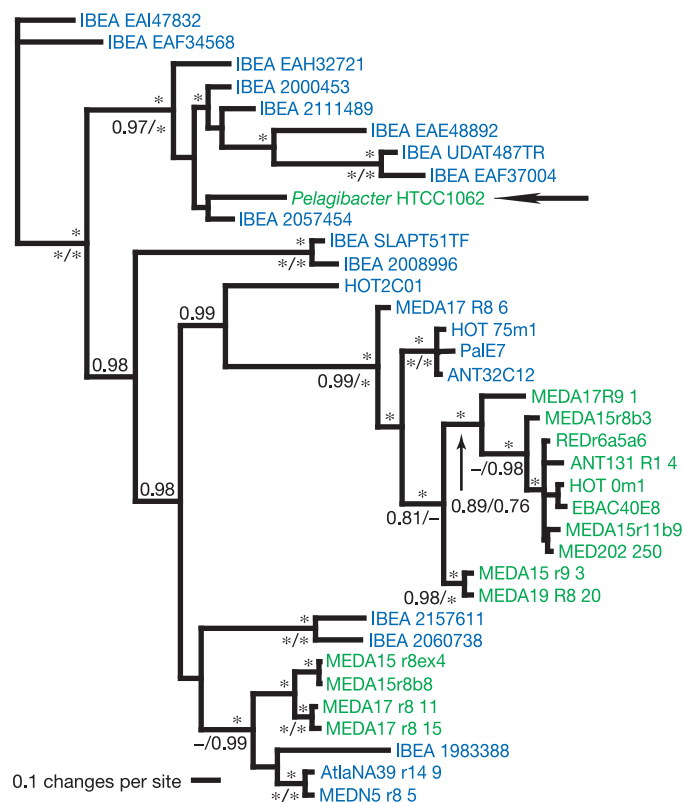


Figure 1 | Phylogenetic relationships between proteorhodopsin amino acid sequences. Shown is the relationship between the *Pelagibacter* strain HTCC1062 proteorhodopsin amino acid sequence and selected representatives of proteorhodopsin genes cloned from seawater DNA. The sequences most similar to *Pelagibacter* proteorhodopsin are included in the tree. Green text indicates proteorhodopsin genes encoding a leucine residue at position 105, blue text indicates a glutamine residue at this position. The tree was rooted with the sequences of rhodopsins from *Gloeobacter violaceus* and *Pyrococcus lunula*. Numbers above nodes are posterior probabilities; numbers below nodes are parsimony bootstrap values and neighbour-joining bootstrap values (separated by a slash). Asterisks indicate a value of 1.0.

¹Department of Microbiology, ²Department of Chemistry, and ³College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. ⁴Diversa Corporation, 4955 Directors Place, San Diego, California 92121-1609, USA. [†]Present addresses: Department of Oceanography, 5N541, Inha University Yungheun-dong, Nam-Gu, Incheon 402-751, Korea (J.-C.C.); Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757-3696, USA (M.D.S.); Hawaii Institute of Marine Biology, SOEST, University of Hawaii, PO Box 1346, Kaneohe, Hawaii 96744, USA (M.S.R.).

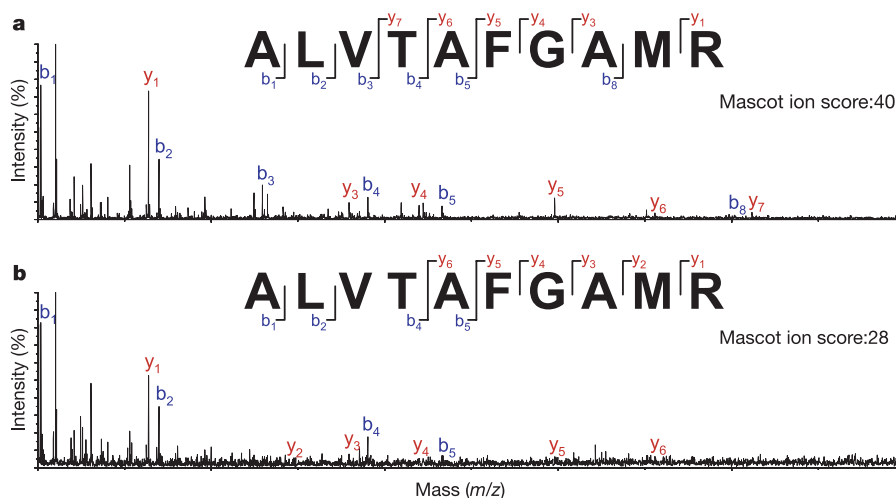


Figure 2 | *Pelagibacter* proteorhodopsin peptide (PR3), detected by tandem MALDI mass spectrometry. **a**, Cultured cells. **b**, Oregon coastal seawater.

the same proteorhodopsin clade, had glutamine residues at the corresponding position, which is typical of proteorhodopsins that absorb blue light.

The genomic region around the HTCC1062 proteorhodopsin gene shows substantial synteny with proteorhodopsin-containing contigs recovered from the Sargasso Sea by environmental DNA sequencing (Supplementary Fig. S1a), suggesting that many of these proteorhodopsin genes might originate from SAR11 strains adapted to life in ocean gyres. Among the Sargasso Sea contigs for which genes downstream of proteorhodopsin were identified, in 49% of the cases the gene encoded a ferridoxin, as observed in the HTCC1062 strain¹⁰. No genes related to carbon fixation were found in the HTCC1062 genome¹⁰. This is consistent with the function of proteorhodopsin proteins, which can supply a transmembrane electrochemical potential but not the reduced nucleotide cofactors needed for carbon dioxide reduction.

We used mass spectrometry to study expression of the proteorhodopsin protein. We used matrix-assisted laser desorption/ionization (MALDI) mass spectrometry, owing to the presence of mostly arginine-containing peptides in the tryptic digest of proteorhodopsin, which are favoured by the MALDI ionization process¹⁴. From membrane-fraction digests of HTCC1062 samples grown in sterilized natural seawater, we identified three proteorhodopsin canonical tryptic peptides with high confidence (Fig. 2a and Supplementary Table S1). The peptides were observed in cultures grown both in diurnal light and in the dark. Notably, two of these proteorhodopsin peptides were also identified in the highly complex protein mixture derived from cells that were collected by filtration from coastal Oregon seawater (Supplementary Table S1 and Fig. 2b). Our mass spectroscopy observations provide strong evidence that native populations of *Pelagibacter* express the proteorhodopsin gene. Other reports have successfully applied mass spectrometry to study the proteome state of natural microbial communities that consist of relatively few species¹⁵. Our successful use of mass spectrometry to detect proteorhodopsin in complex plankton communities can be attributed to the very high abundance of *Pelagibacter* cells in seawater⁸ and to the use of protocols designed to enrich *Pelagibacter* proteorhodopsin protein, by size-fractionating both cells and proteins (see the Supplementary Methods).

We used physical methods to show that the HTCC1062 proteorhodopsin protein has the kinetic characteristics of an ion pump, and also confirmed that it is expressed by cells growing in the light or in darkness. Consistent with predictions from the amino acid sequence, cloned HTCC1062 proteorhodopsin expressed in *Escherichia coli* had an absorption spectrum with a maximum at 530 nm, which is typical of proteorhodopsin genes cloned from ocean surface samples, where

the light field is dominated by green wavelengths¹ (Fig. 3a). Flash-induced absorbance transients from cultures of HTCC1062 grown in the light and in the dark are shown in Fig. 3b, c. The absorption of transient species produced by an excitation flash was monitored as a function of time. The 488-nm absorption transients fit exponential decay models with time constants of 34 ms for dark-grown cells and 13 ms for light-grown cells. Rapid photocycle rates of this order are characteristic of ion pumps rather than sensory rhodopsins, which have slower turnover times. The HTCC1062 proteorhodopsin expressed in *E. coli* showed photocycle rates similar to cells grown in darkness. Suspensions of *E. coli* cells expressing proteorhodopsin produced transient pH decreases when exposed to light, providing further support for the conclusion that this proteorhodopsin is a light-driven proton pump (Supplementary Fig. S2).

We estimated the number of proteorhodopsin molecules per *Pelagibacter* cell by assuming that the ratio between the 633-nm absorbance change 5 ms after the flash, and the absorption of proteorhodopsin at its visible wavelength maximum, is the same in our experiments as cited in the literature². We also assumed an extinction coefficient of $50,000 \text{ M}^{-1} \text{ cm}^{-1}$ for proteorhodopsin at this wavelength. We calculate that *Pelagibacter* cells have on the order of 10,000 proteorhodopsin molecules. Considering the unusually small physical size of these bacteria, this value agrees well with the previously reported estimate of 25,000 proteorhodopsin molecules per bacterioplankton cell from seawater samples².

Pelagibacter cells grew equally well in seawater in the presence and absence of retinal, a cofactor that is required for synthesis of the proteorhodopsin holoenzyme (Fig. 4). An operon was found in the HTCC1062 genome that probably encodes the enzymatic pathway for β -carotene and retinal biosynthesis. In *Halobacterium*, the *crtB*, *crtI* and *crtY* genes catalyse β -carotene synthesis from GGPP (geranylgeranyl pyrophosphate). The *blh* gene is thought to mediate cleavage of β -carotene to retinal¹⁶. We found a *crtIBY* operon in the HTCC1062 genome that was immediately followed by an open reading frame (ORF) with weak but discernable homology (PSI-BLAST expect = 8×10^{-6}) to a *brp* homologue from *Haloarcula marismortui*. This leads us to speculate that *Pelagibacter* has the *blh* gene and is able to produce retinal from β -carotene. The operon also included the *lytB* gene, which encodes a protein that catalyses the conversion of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate into IPP (isopentenyl-diphosphate) and DMAPP (dimethylallyl-diphosphate), which are precursors for geranylgeranyl pyrophosphate synthesis.

The proteorhodopsin photosystem is encoded by a single gene. Despite its simplicity, proteorhodopsin occupies space in the cell membrane, thereby reducing the surface area available for nutrient transporters in an ecosystem in which nanomolar substrate

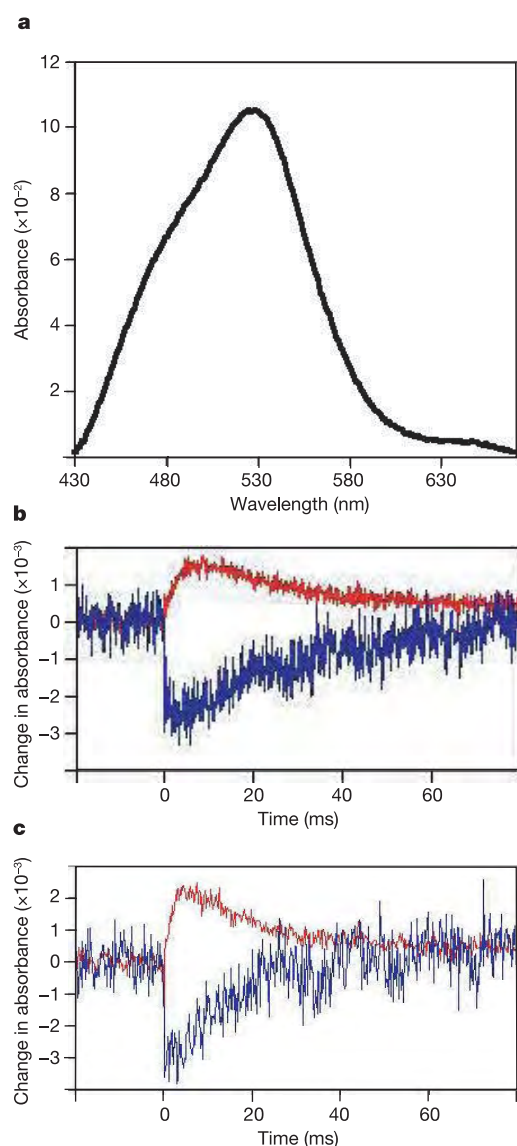


Figure 3 | Spectroscopy. **a**, Absorption spectrum from HTCC1062 proteorhodopsin cloned into an *E. coli* expression vector, showing a characteristic absorption maximum at 530 nm. See main text for experimental details. **b**, **c**, Flash-induced absorption changes in HTCC1062 cells grown on a diurnal light cycle (**b**) or in the dark (**c**), showing short-timescale transients typical of rhodopsin proton pumps. Red, 633-nm absorbance; blue, 488-nm absorbance.

concentrations prevail and nutrient competition is a dominant factor. From the estimated number of proteorhodopsin molecules per *Pelagibacter* cell (10,000), and the measured size of the *Pelagibacter* cytoplasm, we calculate that proteorhodopsin occupies approximately 20% of the inner membrane surface area. Thus, the expression of proteorhodopsin may involve metabolic trade-offs that become fully manifest only when specific conditions that occur periodically in nature are modelled and reproduced.

Notably, we found no consistent differences in growth rate or maximum cell yield in HTCC1062 cultures grown in a diurnal light regime or in complete darkness on natural seawater (Fig. 4). This observation suggests that proteorhodopsins have a subtle role related to the ecological complexities of the ocean surface environment. We speculate that the benefits of proteorhodopsin may be most evident when organic carbon limitation decreases the ability of cells to generate a proton motive force by respiration. This hypothesis cannot be tested at present because the factors that limit the growth

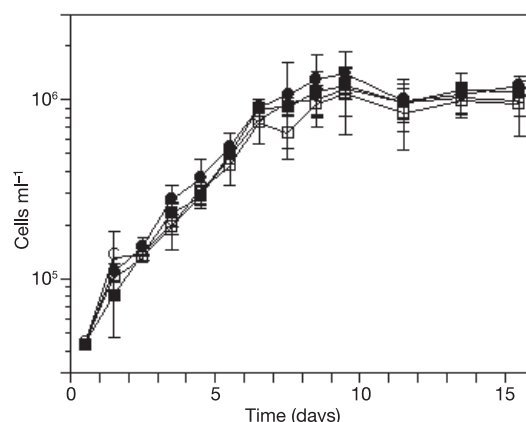


Figure 4 | Growth characteristics of HTCC1062. Bacteria were grown in seawater supplemented with N and P (LNHM) with no added organic carbon, on a diurnal light cycle (open symbols) or in darkness (closed symbols) under high-range light intensity (circles, $680 \mu\text{mol m}^{-2} \text{s}^{-1}$) or middle-range light intensity (squares, $250 \mu\text{mol m}^{-2} \text{s}^{-1}$). Error bars show standard deviation for triplicate experiments. No difference was observed for replicates with and without added retinal (data not shown).

of *Pelagibacter* in seawater remain unidentified. Alternatively, proteorhodopsin may confer a slight advantage in growth rate or biomass yield that has not been resolved experimentally. In some experiments, the cells grew slightly better in the light, but the results varied—possibly as a consequence of variation in the composition of the natural seawater used in different experiments.

Owing to the vast expanse of the ocean habitat, its age and the extraordinary population sizes of bacterioplankton¹⁷, theory predicts that genetic traits conveying subtle differences in fitness should become fixed by selection, yielding strains that are highly adapted to their environment¹⁸. Proteorhodopsin appears to be one of a suite of characteristics that enable bacterioplankton cells to thrive in the competitive, low-nutrient conditions that prevail at sea. Fully understanding the role of this unique photochemistry may ultimately require the delicate reproduction of conditions that occur periodically in the life history of these cells.

METHODS

Cultivation. Cells were cultivated as described in ref. 9, on LNHM (low nutrient heterotrophic medium) medium with the addition of $1 \mu\text{M}$ retinal. Seawater for media was collected with Niskin bottles from station NH5, five miles offshore of Newport, Oregon. For routine cultivation, cells were grown under a diurnal light cycle: cool-white light was supplied at $24 \mu\text{mol m}^{-2} \text{s}^{-1}$ in a 14-h light:10-h dark cycle. Larger banks of cool-white lights were used to increase intensities for the experiments shown in Fig. 4. The light treatments and dark controls for Fig. 4 were incubated in duplicate in the same water baths, and the medium for all treatments was made from the same seawater sample.

Sample collection. Plankton cells were collected at station NH5 from 40 l of seawater (at 10-m depth), pre-filtered with an $0.8\text{-}\mu\text{m}$ filter (Supor) and concentrated by tangential flow filtration with a Millipore Pellicon II Mini system equipped with a 30-kDa regenerated cellulose filter. Concentrated cells were pelleted by centrifuging in a Beckman J2-21 centrifuge using a JA-20 rotor at 48,400g and a temperature of 4°C . Cells were resuspended in a minimal volume of seawater and stored at -20°C until analysis.

Gene cloning. The proteorhodopsin gene was amplified from isolate HTCC1062 by polymerase chain reaction using primers 5'-ACCATGGGTAAAAA CTAAAATTGTTTGC-3' and 5'-CTTAGCTCTACCAGGTGAGA-3'. Cloning and expression were performed as described in ref. 1.

Phylogenetic analysis. Sequences were aligned using Clustal W. Bayesian analyses were conducted using MRBAYES 3.0 (ref. 19); searches were conducted for a total of 1,000,000 generations with phylogenetic trees sampled every 100 generations. Monte Carlo Markov chains were calculated using an integration of ten fixed amino-acid rate matrices. Out of 10,000 resulting trees, the initial 1,000 generations were identified as preceding the convergence of likelihoods and excluded from post-run analyses (burn-in). A majority-rule consensus tree with

averaged branch lengths and posterior probabilities was calculated from the remaining 9,000 trees (Fig. 1). Maximum parsimony and neighbour-joining analyses were conducted using PAUP* 4.0b10, with the following settings: 100 replicates of random sequence addition, TBR branch swapping, and MulTrees in effect. The relative support for the resulting trees was determined by 1,000 bootstrap replications with the same search options as previously described.

Mass spectrometry. Mass spectrometry (MS) methods are described elsewhere²⁰. Briefly, cultured cell pellets were washed and lysed, and membrane material was collected by centrifugation and then dissolved with the detergent dodecyl maltoside. Proteins were then digested in-solution using trypsin. Total protein from lysates of environmental cell pellets was separated by one-dimensional SDS-PAGE. Gel lanes were cut into six pieces, and in-gel tryptic digestions were performed. Chromatography was performed using a Waters CapLC system with a 0.32-mm (inner diameter) symmetry column packed with 5- μ m C18 particles, and the eluate was spotted onto MALDI plates using a MALDIprep spotter (Waters Corp). MALDI-MS and tandem MS were performed on an Applied Biosystems 4700 proteomics analyser. Peptides were identified using the program Mascot²¹.

Spectroscopy. Attenuance spectra of proteorhodopsin were acquired from standard 1-cm cuvettes on a Cary 300 spectrophotometer (Varian Instruments), using intact *E. coli* cells expressing the cloned proteorhodopsin protein. The absorption spectrum depicted in Fig. 3a was obtained by subtracting a cubic baseline, determined by fitting the attenuance values from 350–450 nm and from 625–800 nm, which represents scattering of the cell suspension and absorption of *E. coli* exclusive of proteorhodopsin. Transient absorption spectra (Fig. 3b, c) of HTCC1062 cells in 1-cm pathlength fluorometer sub-microcuvettes (16.160F-Q-10/Z20, Starna Cells) were acquired on a laboratory-constructed flash-photolysis apparatus. Each flash-induced transient absorption trace was obtained by averaging 32 acquisitions, one every 25 s. The excitation flash (10 mJ pulse⁻¹) at 532 nm was supplied by a Nd:YAG laser (Moletron MY-34-10, Coherent). The probe beams were supplied by a helium-neon laser (633 nm; Hughes 5000, Melles Griot) and an argon laser (488 nm; Omnicrome 532, Melles Griot). These wavelengths were chosen because for probe wavelengths shorter than the transient isosbestic point (at about 550 nm), the transient absorption is expected to be negative, whereas the converse is expected for wavelengths longer than the isosbestic wavelength² (as observed in Fig. 3b, c). Polarizations of the laser beams were set to 'magic angle' orientations (valid because the transient absorption is small).

Received 29 November 2004; accepted 11 July 2005.

- Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
- Sabehi, G., Béjà, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6**, 903–910 (2004).
- Man-Aharonovich, D. *et al.* Characterization of RS29, a blue-green proteorhodopsin variant from the Red Sea. *Photochem. Photobiol. Sci.* **3**, 459–462 (2004).
- de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100**, 12830–12835 (2003).
- Sabehi, G. *et al.* Novel Proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* **5**, 842–849 (2003).
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810 (2002).
- Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
- Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
- Field, K. G. *et al.* Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **61**, 63–70 (1997).
- Wang, W. W., Sineshchekov, O. A., Spudich, E. N. & Spudich, J. L. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* **278**, 33985–33991 (2003).
- Man, D. *et al.* Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
- Stapels, M. D. & Barofsky, D. F. Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. *Anal. Chem.* **76**, 5423–5430 (2004).
- Ram, R. J. *et al.* Community proteomics of a natural microbial biofilm. *Science* **308**, 1915–1920 (2005).
- Peck, R. F., Johnson, E. A. & Krebs, M. P. Identification of a lycopene β -cyclase required for bacteriorhodopsin biogenesis in the archaeon *Halobacterium salinarum*. *J. Bacteriol.* **184**, 2889–2897 (2002).
- Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Stapels, M. D., Cho, J. C., Giovannoni, S. J. & Barofsky, D. F. Proteomic analysis of novel marine bacteria using MALDI and ESI mass spectrometry. *J. Biomol. Tech.* **15**, 191–198 (2004).
- Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Nibler and the chemistry department at Oregon State University for the use of the Nd:YAG laser and student laser laboratory facilities; W. Hetherington, R. Letelier, B. Geller and O. Béjà for helpful discussions; and E. L. Barofsky for her assistance with MALDI mass spectrometry. This research was supported by the National Science Foundation, Diversa Corporation and the National Institute of Environmental Health Sciences.

Author Contributions S.J.G. led the genome sequencing project, provided the bioinformatics analyses and was the primary writer. L.B. and E.J.M. led the DNA sequencing team at Diversa Corporation. J.-C.C., L.J.W. and H.J.T. provided the growth data. M.D.S. and D.F.B. provided the mass spectrometry analysis. R.D. and S.L. performed the light spectroscopy experiments. K.L.V. cloned the proteorhodopsin gene and showed that it was a light-dependent proton pump, with the assistance of R.D. M.S.R. isolated the *Pelagibacter*.

Author Information The HTCC1062 proteorhodopsin gene sequence has been deposited in GenBank under accession number CP000084. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.G. (steve.giovannoni@oregonstate.edu).

LETTERS

Photosynthesis genes in marine viruses yield proteins during host infection

Debbie Lindell¹, Jacob D. Jaffe^{2†}, Zackary I. Johnson^{1†}, George M. Church² & Sallie W. Chisholm^{1,3}

Cyanobacteria, and the viruses (phages) that infect them, are significant contributors to the oceanic 'gene pool'^{1,2}. This pool is dynamic, and the transfer of genetic material between hosts and their phages^{3–6} probably influences the genetic and functional diversity of both. For example, photosynthesis genes of cyanobacterial origin have been found in phages that infect *Prochlorococcus*^{5,7} and *Synechococcus*^{8,9}, the numerically dominant phototrophs in ocean ecosystems. These genes include *psbA*, which encodes the photosystem II core reaction centre protein D1, and high-light-inducible (*hli*) genes. Here we show that phage *psbA* and *hli* genes are expressed during infection of *Prochlorococcus* and are co-transcribed with essential phage capsid genes, and that the amount of phage D1 protein increases steadily over the infective period. We also show that the expression of host photosynthesis genes declines over the course of infection and that replication of the phage genome is a function of photosynthesis. We thus propose that the phage genes are functional in photosynthesis and that they may be increasing phage fitness by supplementing the host production of these proteins.

Photosynthesis in cyanobacteria, algae and plants requires two photosystems (denoted PSI and PSII). The D1 and D2 proteins (encoded by *psbA* and *psbD*, respectively) form a heterodimer in the reaction centre of PSII and bind the components required for photochemistry. The D1 protein is turned over rapidly owing to light-induced damage¹⁰; thus, its *de novo* synthesis is required for sustained photosynthesis¹⁰. High-light-inducible proteins (HLIPs) protect the photosynthetic apparatus from photodamage by dissipating excess light energy¹¹.

Numerous cyanophages contain photosynthesis genes (*psbA* and at least one other)^{5,7–9} with highly conserved amino acid sequences^{7,12}, suggesting that they encode functional proteins that may be involved in maintaining host photosynthesis during infection. Here we used *Prochlorococcus* MED4 and the podovirus P-SSP7 (a T7-like phage⁵) as a model system to begin exploring this hypothesis. We considered that if these genes are involved in host photosynthesis, then the amount of phage production might be dependent on photosynthetic performance and, conversely, host photosynthesis might be compromised by phage infection. We examined the first part of this hypothesis by inhibiting photosynthesis with darkness or DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea), an inhibitor of electron flow from PSII to PSI, which led to a respective four- or twofold reduction in replication of the phage genome (Fig. 1a). Thus, as in other systems^{13–15}, continued photosynthesis is necessary for maximal phage replication. To determine the converse, that is, whether host photosynthesis is influenced by phage infection, we measured PSII photochemical conversion efficiency (F_v/F_m) and functional cross-sectional area

(σ_{PSII}) during the 8-h latent period before lysis. The former decreased only slightly, whereas the latter was constant throughout this period (Fig. 1b, c), indicating that phage infection does not lead to a marked decline in PSII performance, as occurs in some^{14–16}, but not other^{17–19} photosynthetic host–virus systems.

Thus, continued photosynthesis is required for maximum phage production in our system. Moreover, photosynthesis is sustained during infection when a decline in the transcription and translation of host genes might be expected, suggesting that the expression of phage photosynthesis genes might be supplementing host metabolism. To address this hypothesis, we determined whether these phage genes are expressed and, if so, how their expression relates to that of the homologous host genes. Using probes specific for the phage and host *psbA* and *hli* genes, we found that both of the phage genes were transcribed (Fig. 2a, b and Supplementary Fig. 1). Using polymerase chain reaction with reverse transcription (RT–PCR), we determined

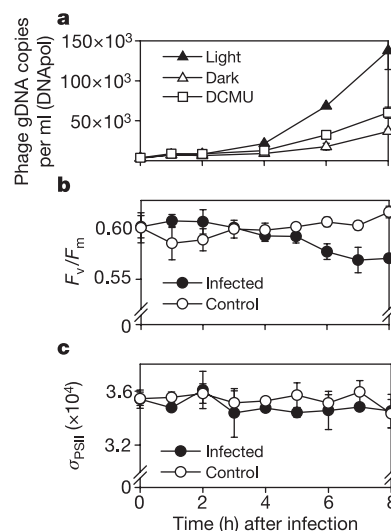


Figure 1 | Photosynthesis and phage infection. **a**, Replication of the phage genome was assessed by quantifying the phage gene encoding DNA polymerase in host cells kept in the light, transferred to the dark or treated with DCMU while in the light. Significantly lower ($P < 0.05$) phage DNA was detected in dark- and DCMU-treated cells after 4 h. **b**, **c**, PSII photochemical conversion efficiency (F_v/F_m ; **b**) and PSII functional absorption cross-sectional area (σ_{PSII} ; **c**) in infected and control cells. The absorption cross-section remained constant, although there was a 10% decline in PSII conversion efficiency. Error bars indicate the s.d. from biological replicates.

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present addresses: The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02141, USA (J.D.J.); Department of Oceanography, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA (Z.I.J.).

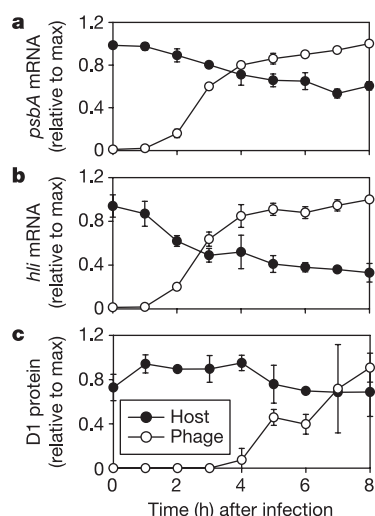


Figure 2 | Expression of phage and host photosynthesis genes. Temporal patterns of *psbA* mRNA (a), *hli12* mRNA (b) and D1 peptides (c). Expression during infection was normalized to expression in uninfected control cells and is presented relative to the maximal level for each individual gene. Data in a and b were obtained from microarray analyses. See Supplementary Fig. 1 for confirmation of the results shown in a by RT-PCR. Host D1 peptides were significantly lower in the second half of the latent period (5–8 h) than in the first half (0–4 h); $P < 0.01$. Error bars indicate the s.d. from biological replicates.

that phage *psbA* mRNA made up 50% of the total (phage plus host) *psbA* transcripts by 7–8 h after infection. By 4–5 h after infection, mRNA from the single host *psbA* gene had dropped to 50–60% of maximal levels (Fig. 2a and Supplementary Fig. 1). Transcription of 16 out of 22 of the host *hli* genes also declined significantly during infection (*hli12* is shown as a representative of this gene family in Fig. 2b). Because the maximal level of first-round infection achieved so far with this host–phage system is 50% (D.L. and S.W.C.,

unpublished data), the presence of only 50–60% of host *psbA* and *hli* transcripts relative to the control suggests that transcription of these genes had almost ceased by 4 h after infection and that nearly all *psbA* transcripts in infected cells were transcribed from the phage genome. The decline in host gene expression was not specific to photosynthesis genes but was part of a general reduction in host transcription subsequent to infection (D.L. and S.W.C., unpublished data).

The decline in host *psbA* transcription should cause a reduction in translation of the D1 protein and, because this protein is turned over rapidly¹⁰, a decrease in host D1 titre. To verify this, we identified and quantified peptides specific for the host and phage D1 proteins (Fig. 3). The host D1 protein declined during infection to roughly 75% of maximal levels (Fig. 2c), reflecting a decrease of 50% in infected cells. Accompanying this reduction was a steady increase in the homologous protein encoded by the phage (Fig. 2c), which made up about 10% of total D1 in infected cells by the end of the latent period.

Although these results are consistent with our hypothesis that the expression of phage D1 protein helps to bolster host photosynthesis, the amount of phage D1 did not quantitatively compensate for the loss in host D1, even though PSII efficiency in the hosts declined very little during infection (Fig. 1b). This suggests that additional factors may be involved in maintaining host photosynthesis. For example, phage HLIPIs, detected from 4 h after infection (data not shown), may reduce photosystem damage as well as being involved in the re-assembly of PSII²⁰. Furthermore, phage D1 may be more efficient than host D1 during infection⁶. If photosynthetic antennae are shared among reaction centres, which is consistent with other studies²¹, then fewer, more efficient functional reaction centres could lead to an increase in the functional absorption cross-section; however, this was not observed (Fig. 1c). Thus, although we are left with an imperfect balance sheet for the host and phage D1 protein, the inverse temporal expression of host and phage photosynthesis genes is striking and suggests that there is a functional interdependency.

A fitness advantage conferred by phage photosynthesis genes not

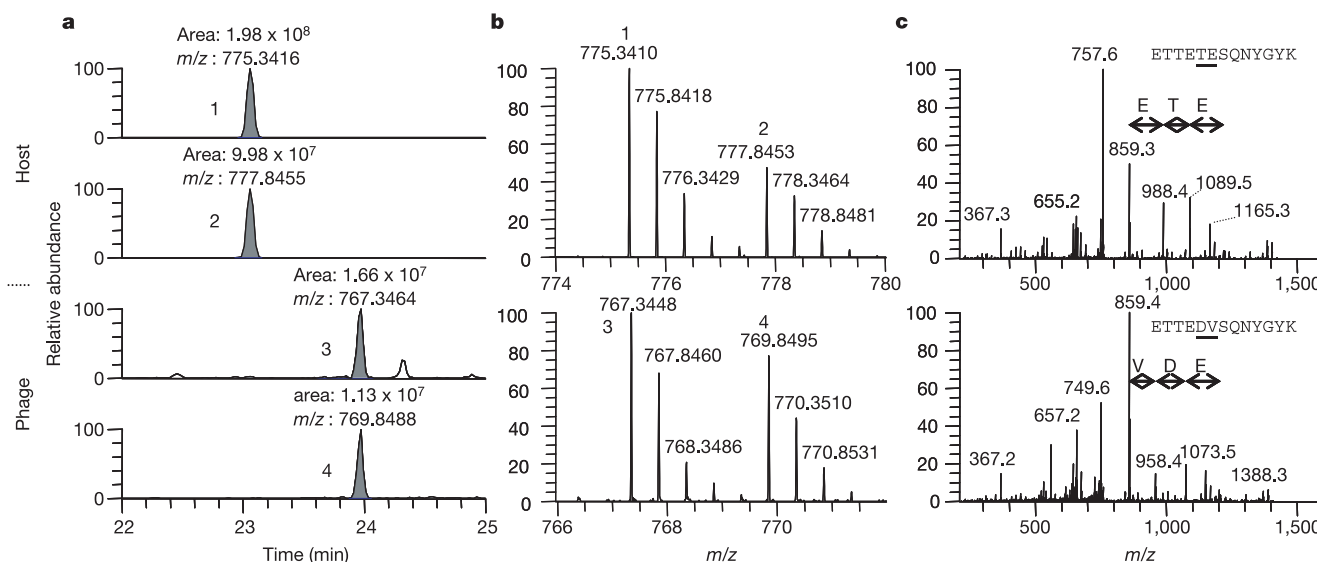


Figure 3 | Analysis of host and phage peptides. a, Extracted ion chromatograms of endogenous host (1) and isotopically labelled synthetic host (2) peptides, and endogenous phage (3) and isotopically labelled synthetic phage (4) peptides. Endogenous and corresponding synthetic peptides co-elute, whereas host and phage peptides have different retention times and m/z values. The area under the peaks of the endogenous peptides was used for quantification. b, Mass spectra taken at the apex of the chromatographic peaks show that host and phage peptides have different

m/z values, as do the endogenous and isotopically labelled synthetic peptide pairs. Note that the threonine residue in the third position of the synthetic host and phage peptides was uniformly labelled with ^{15}N and ^{13}C isotopes, adding 5 Da to the mass of the peptides. c, Collision-induced dissociation spectra facilitate sequence identification of the host and phage peptides. The amino acid sequence in the partial peptide annotation is reversed because the y -ion series is shown. The full sequence of each peptide is shown with each spectrum.

only would explain their presence in cyanophage genomes^{7,9}, but also supports the modular theory of phage evolution²², in which phages evolve through the step-wise acquisition of genes from diverse sources. According to this theory, acquired genes are initially expressed autonomously and are integrated into the phage life cycle if they provide a fitness advantage. The photosynthesis genes in cyanophage originate from cyanobacteria^{7,9,12}, and these phage genomes also contain bacterial, and even archaeal and eukaryotic genes^{5,6}. Notably, the *psbA* and *psbD* genes in two *Prochlorococcus* myoviruses⁷ have putative promoter and transcriptional terminators flanking the genes, suggesting that they are autonomously expressed. By contrast, photosynthesis genes in the phage used here have overlapping start and stop codons⁷ and are co-transcribed with the essential, highly expressed phage capsid genes surrounding the photosynthesis genes (Fig. 4), suggesting that they have become an integral part of the phage genome. Thus, the proposed fitness advantage conferred by these photosynthesis genes, and other genes acquired from their hosts^{5,6}, may be a forerunner to their becoming *bona fide* members of the cyanophage gene pool.

Although we have not proved that phage photosynthesis gene products are participating in host photosynthesis, we favour this hypothesis, first, because photosynthesis continues during infection despite the decline in expression of host photosynthesis genes; second, because maximal phage DNA replication is dependent on this sustained photosynthesis; and last, because the high conservation of amino acid sequences^{7,12} in the phage proteins suggests that these proteins are functioning in the same role as the host proteins. Nonetheless, we cannot rule out alternative functions. One could argue that D1, an efficient manganese-binding protein¹⁰, is obtaining this metal ion for phage enzymatic reactions. Or perhaps phage D1, like the HLP, is involved in dissipating excess light energy without being actively involved in photosynthesis. Obviously, the next stages in testing the 'same function hypothesis' are to see whether phage D1 localizes to the host PSII complex and to study the behaviour of this phage-host system using phage in which the photosynthesis genes have been inactivated.

The dynamic nature of the oceanic gene pool has led to the genetic

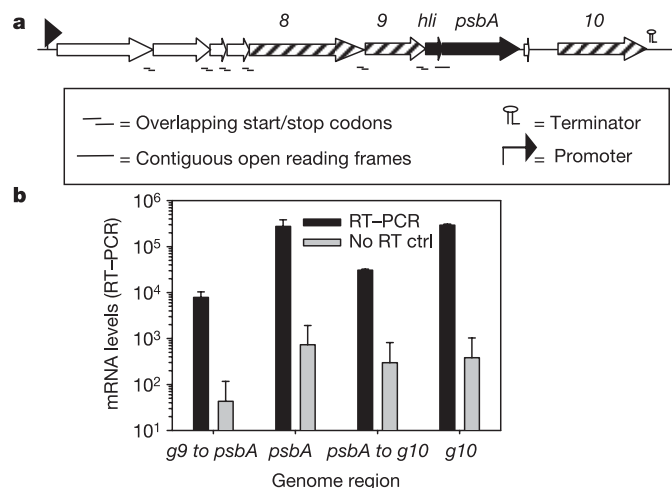


Figure 4 | Co-transcription of phage photosynthesis and capsid genes. **a**, Proposed operon of the photosynthesis gene region determined bioinformatically⁷. **b**, Transcript levels of regions internal to the *psbA* and *g10* genes, and regions spanning from *psbA* to the capsid genes upstream (*g9* to *psbA*) and downstream (*psbA* to *g10*) of *psbA* at 4 h after infection. *g9* encodes capsid assembly protein, *g10* encodes major capsid protein. Amplification from mRNA ('RT-PCR') was more than two orders of magnitude greater than from genomic DNA ('no RT control'). The lower transcript abundance across gene boundaries may be due to posttranscriptional processing or additional autonomous transcription of *psbA* and *g10*. Error bars indicate the s.d. from biological replicates.

diversification of donor and recipient genomes⁷ and has apparently guided their functional diversification as well. If phage photosynthesis proteins do indeed function in host photosynthesis, this is a striking example of the interaction of proteins encoded from two distinct genomes in a single metabolic complex. Furthermore, the abundance of cyanobacteria²³ and their phages^{16,24} in the oceans suggests that phage photosynthesis proteins have a small but significant role in the conversion of light to chemical energy on a global scale.

METHODS

Experimental conditions. *Prochlorococcus* MED4 was grown at 21 °C under continuous cool white light (25 μmol photon m⁻² s⁻¹) in Sargasso seawater Pro99 medium²⁵ amended with 10 mM HEPES (pH 7.5) and 12 mM sodium bicarbonate. Cells were concentrated to 10⁸ cells per ml by centrifugation, and triplicate cultures were infected with 3 × 10⁸ infective phages per ml for a multiplicity of infection (MOI) of 3 (except for the experiment shown in Fig. 1a; see below). Controls were amended with filter-sterilized spent medium. Cells for RNA and protein analyses were collected by centrifugation (12,400 g for 15 min at 20 °C), resuspended in buffer (200 mM sucrose, 10 mM sodium acetate, 5 mM EDTA; pH 5.2), snap frozen in liquid nitrogen and stored at -80 °C. Sample handling took 30 min. For the experiment shown in Fig. 1a, *Prochlorococcus* at 10⁸ cells per ml were exposed to 10⁷ infective phages per ml for an MOI of 0.1. After 1 h in the light to allow adsorption, cultures were diluted 4,000-fold and incubated in the light or dark, or in the light with 50 μM DCMU. We determined the MOI of phage stocks by the most probable number assay.

Quantitative detection of phage genomic DNA. *Prochlorococcus* cells were collected on 0.2-μm pore-sized polycarbonate filters (Osmonics), washed with sterile seawater followed by 3 ml of preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl; pH 8) and frozen at -80 °C. We prepared DNA by a heat lysis method²⁶. Phage genomic DNA was quantified by real-time PCR (see below), targeting the phage DNA polymerase gene. Primer sequences are given in Supplementary Table 1.

Photosynthesis measurements. A background irradiance gradient single-turnover fluorometer (BIG-STf) was used to measure the photosynthetic conversion efficiency (F_v/F_m) and functional absorption cross-section area (σ_{PSII}) of PSII, which measures the ability of PSII to absorb photons from antennae complexes²⁷. Triplicate samples were dark acclimated for 15–30 min before single-turnover fluorescence induction curve measurements. F_v/F_m and σ_{PSII} were estimated by fitting standard models²⁸ to the data to determine F_o (initial fluorescence), F_m (maximal fluorescence), F_v ($F_m - F_o$) and σ_{PSII} .

RNA extraction and transcript analysis. Total RNA was extracted by a mirVana RNA isolation kit (Ambion). We removed DNA by a Turbo DNA-free kit (Ambion). For microarray analysis, RNA was concentrated by ethanol precipitation, and 2 μg of total RNA was labelled and hybridized to custom-made MD4-9313 arrays (Affymetrix) using the standard Affymetrix protocol for *Escherichia coli* (<http://www.affymetrix.com/technology/index.affx>). This array contains probe sets for genes and intergenic regions of both host and phage. Standard affymetrix procedures were used for probe design and construction of the array. For RT-PCR, total RNA (0.5–10 ng) was reverse transcribed with gene-specific primers and 100 U of SuperScript II (Invitrogen) in the presence of 200 U of SuperscriptN (Ambion). Triplicate real-time PCR reactions were done with a QuantiTect SYBR Green PCR kit (Qiagen) and primers at 0.3–1.0 μM. After 15 min at 95 °C, 40 cycles of denaturation (95 °C, 15 s), annealing (56 °C, 30 s) and elongation (72 °C, 30 s) were run on an DNA Engine Opticon (MJ Research), which was followed by 5 min at 72 °C and melt curve analysis. Incorporation of SYBR stain into double-stranded DNA was determined subsequent to elongation steps. Standard curves were generated with genomic DNA from *Prochlorococcus* or P-SSP7 phage particles. Primer sequences are given in Supplementary Table 1.

Protein preparation and analysis. Cells were lysed in 3 M urea, 0.05% SDS and 50 mM Tris-HCl (pH 8). Proteins were digested with sequencing grade trypsin (Promega) at a protein to trypsin ratio of 137.5:1, reduced with 10 mM dithiothreitol, alkylated with 50 mM iodoacetamide and acidified to pH < 3. Total protein was purified by solid-phase extraction using Oasis MCX (Waters), concentrated by vacuum centrifugation, resuspended in 0.1% formic acid and purified by solid-phase extraction using Oasis HLB (Waters). Peptides eluted with 70% acetonitrile were concentrated by vacuum centrifugation to dryness and resuspended in 5% acetonitrile and 5% formic acid. Tryptic digestions yielded peptides suitable for differentiation between the host and phage isoforms of D1. Peptides with the sequences NH₂-ETTETESQNYGYK-COOH and NH₂-ETTEDVSQNYGYK-COOH were observed by liquid chromatography mass spectrometry (LC-MS; Fig. 3) and used as surrogates for the host and phage

D1 proteins, respectively. Synthesized stable isotope-containing variants of these peptides were used as mass spectrometric standards to verify the elution time and tandem MS fragmentation patterns of the host and phage peptides (Fig. 3). D1 peptides were identified and quantified by reversed-phase LC-MS as described²⁹ from duplicate injections of 5.7 µg of total peptides using a hybrid linear ion trap, Fourier transform ion cyclotron resonance mass spectrometer (ThermoElectron) with resolution set to 100,000 and a mass accuracy of ± 8 p.p.m. Extracted ion chromatograms were generated by monitoring the mass/charge (*m/z*) values of 775.3363 ± 0.006 and 767.3388 ± 0.006 for the host and phage peptides, respectively. We used the area under the peaks from XCalibur Software for quantification (see also <http://arep.med.harvard.edu/mapquant.html> for an alternative method).

Received 27 June; accepted 27 July 2005.

Published online 12 October 2005.

- Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brussow, H. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**, 417–424 (2003).
- Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424**, 1037–1042 (2003).
- Sullivan, M. B., Coleman, M. L., Weigle, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biology* **3**, e144 (2005).
- Mann, N. H. *et al.* The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus*. *J. Bacteriol.* **187**, 3188–3200 (2005).
- Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018 (2004).
- Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
- Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc. Natl Acad. Sci. USA* **101**, 11007–11012 (2004).
- Adir, N., Zer, H., Shochat, S. & Ohad, I. Photoinhibition—a historical perspective. *Photosynth. Res.* **76**, 343–370 (2003).
- Havaux, M., Guedeney, G., He, Q. & Grossman, A. R. Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll *a/b*-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochim. Biophys. Acta* **1557**, 21–33 (2003).
- Zeidner, G. *et al.* Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ. Microbiol.* **7**, 1505–1513 (2005).
- Adolph, K. W. & Haskelkorn, R. Photosynthesis and the development of blue-green algal virus N-1. *Virology* **47**, 370–374 (1972).
- MacKenzie, J. J. & Haselkorn, R. Photosynthesis and the development of blue-green algal virus SM-1. *Virology* **49**, 517–521 (1972).
- Sherman, L. A. Infection of *Synechococcus cedrorum* by the cyanophage AS-1M. *Virology* **71**, 199–206 (1976).
- Suttle, C. A. & Chan, A. M. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*—abundance, morphology, cross-infectivity and growth-characteristics. *Mar. Ecol. Prog. Ser.* **92**, 99–109 (1993).
- Ginzburg, D., Padan, E. & Shilo, M. Effect of cyanophage infection on CO₂ photoassimilation in *Plectonema boryanum*. *J. Virol.* **2**, 695–701 (1968).
- Rahoutei, J., Garcia-Luque, I. & Baron, M. Inhibition of photosynthesis by viral infection: effect on PSII structure and function. *Physiol. Plant.* **110**, 286–292 (2000).
- Arias, M. C., Lenardon, S. & Taleisnik, E. Carbon metabolism alterations in sunflower plants infected with the sunflower chlorotic mottle virus. *J. Phytopath.* **151**, 267–273 (2003).
- Xu, H., Vavilin, D., Funk, C. & Vermaas, W. Multiple deletions of small Cab-like proteins in the cyanobacterium *Synechocystis* sp. PCC 6803—consequences for pigment biosynthesis and accumulation. *J. Biol. Chem.* **279**, 27971–27979 (2004).
- Bruyant, F. *et al.* Diel variations in the photosynthetic parameters of *Prochlorococcus* strain PCC 9511: combined effects of light and cell cycle. *Limnol. Oceanogr.* **50**, 850–863 (2005).
- Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8**, 504–508 (2000).
- Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
- Moore, L. R., Post, A. F., Roca, G. & Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47**, 989–996 (2002).
- Zinser, E. R. *et al.* *Prochlorococcus* ecotype abundance in the North Atlantic Ocean revealed by an improved quantitative PCR method. *Appl. Environ. Microbiol.* (in the press).
- Johnson, Z. I. Development and application of the background irradiance gradient—single turnover fluorometer (BIG-STf). *Mar. Ecol. Prog. Ser.* **283**, 73–80 (2004).
- Kolber, Z. S., Prasil, O. & Falkowski, P. G. Measurements of variable chlorophyll fluorescence using fast repetition rate techniques—defining methodology and experimental protocols. *Biochim. Biophys. Acta* **1367**, 88–106 (1998).
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Rector and R. Steen for doing the Affymetrix GeneChip experiments; C. Steglich, M. Sullivan, M. Coleman, and E. Zinser for discussions; and M. Sullivan for comments on the manuscript. This research was supported by grants from the National Science Foundation (to S.W.C.), the Gordon and Betty Moore Foundation's Program in Marine Microbiology (to S.W.C.), and the Department of Energy Genomes to Life Program (to S.W.C. and G.M.C.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu).

LETTERS

Algae acquire vitamin B₁₂ through a symbiotic relationship with bacteria

Martin T. Croft¹, Andrew D. Lawrence², Evelyne Raux-Deery², Martin J. Warren² & Alison G. Smith¹

Vitamin B₁₂ (cobalamin) was identified nearly 80 years ago as the anti-pernicious anaemia factor in liver¹, and its importance in human health and disease has resulted in much work on its uptake², cellular transport³ and utilization⁴. Plants do not contain cobalamin because they have no cobalamin-dependent enzymes. Deficiencies are therefore common in strict vegetarians⁵, and in the elderly, who are susceptible to an autoimmune disorder that prevents its efficient uptake⁶. In contrast, many algae are rich in vitamin B₁₂, with some species, such as *Porphyra yezoensis* (Nori), containing as much cobalamin as liver⁷. Despite this, the role of the cofactor in algal metabolism remains unknown, as does the source of the vitamin for these organisms. A survey of 326 algal species revealed that 171 species require exogenous vitamin B₁₂ for growth, implying that more than half of the algal kingdom are cobalamin auxotrophs. Here we show that the role of vitamin B₁₂ in algal metabolism is primarily as a cofactor for vitamin B₁₂-dependent methionine synthase, and that cobalamin auxotrophy has arisen numerous times throughout evolution, probably owing to the loss of the vitamin B₁₂-independent form of the enzyme. The source of cobalamin seems to be bacteria, indicating an important and unsuspected symbiosis.

Vitamin B₁₂ is one of nature's most complex metabolites, requiring at least 19 separate enzymatic steps for its synthesis from uroporphyrinogen III, the common precursor of all tetrapyrroles including haem and chlorophyll. These enzymes have been characterized in prokaryotic organisms only⁸. Many marine algae are known to be extremely rich in vitamin B₁₂, but the source remains a matter of controversy. Although there have been reports that some algae are able to synthesize cobalamin *de novo*⁹, many other algae have been found to require exogenous cobalamin for growth in culture¹⁰, implying that they are unable to make it themselves. The vitamin B₁₂ requirements of different species of algae grown in axenic culture were evaluated in several early projects, but there has been no systematic examination across the algal kingdom. We therefore compiled the data in the literature for more than 300 species of algae (Supplementary Table S1). To assess the validity of the data, we grew a number of representative species from each phylum (highlighted in Supplementary Table S1) in minimal medium and tested their vitamin B₁₂ requirements. In every case our results coincided with those previously published. Table 1 presents a summary of our analysis by algal group. Of the 326 species surveyed, over half require vitamin B₁₂ for growth, demonstrating that these organisms are not true autotrophs. Furthermore, this requirement shows no relationship to established algal lineages: all the phyla contain species that require the vitamin and species that do not. This pattern is also mirrored within individual genera such as *Lobomonas* in the Chlorophyta, *Peridinium* in the Dinophyta, *Pavlova* in the Haptophyta and *Thalassiosira* in the Heterokontophyta. This suggests that vitamin B₁₂ auxotrophy has arisen independently numerous times during evolution, and that these algae cannot synthesize vitamin B₁₂ *de novo*. This

poses the question, do the remaining algae synthesize the vitamin, or have they dispensed with it altogether like higher plants?

Recently, the genome sequences of three algae have been released: *Chlamydomonas reinhardtii* (see <http://www.jgi.doe.gov/>), a green alga (Chlorophyta); *Cyanidioschyzon merolae*¹¹, a red alga (Rhodophyta); and the diatom (Heterokontophyta) *Thalassiosira pseudonana*¹². Our own growth experiments confirm that the first two organisms do not require vitamin B₁₂, whereas *T. pseudonana* shows a clear dependence on it (Supplementary Table S1). *C. reinhardtii* has been proposed to synthesize the vitamin *de novo*, because when the organism was transferred from a medium containing vitamin B₁₂ to unsupplemented medium, the cells still contained traces of the vitamin¹³. However, a search of the *C. reinhardtii* genome found no genes with sequence similarity to those encoding vitamin B₁₂ biosynthetic enzymes. Similarly, there are no biosynthetic genes in the *C. merolae* genome. *T. pseudonana*, on the other hand, contains a gene with sequence similarity to *cbiP*, which encodes cobyrinic acid *a,c*-diamide synthase. However, when we carried out a polymerase chain reaction with reverse transcription (RT-PCR) analysis, we failed to detect transcripts of this gene (Fig. 1a), which suggests that it is not expressed. Furthermore, this organism contains no genes for any of the other 18 enzymes specific for vitamin B₁₂ biosynthesis.

It is conceivable that cobalamin is synthesized through an alternative pathway to that found in bacteria. To test this, five vitamin B₁₂-independent algae (Supplementary Table S1), including *C. reinhardtii*, were grown for at least five subcultures in unsupplemented medium and assayed for the presence of vitamin B₁₂ using a bioassay with a detection limit of 1 ng ml⁻¹. In every case, there was no measurable cobalamin in cell extracts. The most likely explanation for the earlier report of vitamin B₁₂ biosynthesis in *C. reinhardtii*¹³ is that the cobalamin present in the cells had been taken up from the supplemented medium. Coupled with the data from the three published genomes, it can be concluded that the pathway for vitamin

Table 1 | A summary of the vitamin B₁₂ requirements of different algal phyla

Phylum	Species surveyed	Require B ₁₂	Do not require B ₁₂
Chlorophyta	154	49	105
Glaucocystophyta	1	1	0
Rhodophyta	13	12	1
Cryptophyta	8	7	1
Dinophyta	30	26	4
Euglenophyta	15	13	2
Haptophyta	22	14	8
Heterokontophyta	83	49	34
Total	326	171	155

Data from the individual species detailed in Supplementary Table S1 are compiled under the different algal groups. The first three groups have simple plastids, resulting from primary endosymbiosis, whereas the remaining groups have complex plastids, due to secondary and tertiary endosymbioses²⁸.

¹Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. ²Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK.

B₁₂ biosynthesis is not present in algae, supporting the generally held view that the ability to synthesize the vitamin did not make the prokaryotic to eukaryotic transition⁸. Cyanobacteria (the ancestors of chloroplasts) and the α -subgroup of proteobacteria (which gave rise to mitochondria) contain members that have the capacity to synthesize vitamin B₁₂, as do the Archaea. However, none of the sequenced genomes from lower eukaryotes (including the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, and the protista *Entamoeba histolytica* and *Dictyostelium discoideum*) contains genes for a functional vitamin B₁₂ biosynthetic pathway.

Vitamin B₁₂ auxotrophy in individual algal species is thus likely to have arisen because the cofactor became essential for their metabolism, rather than from the loss of a functional biosynthetic pathway. To address this further, we investigated the role of the biological form of vitamin B₁₂ in algae. We searched the three algal genomes for genes that encode vitamin B₁₂-dependent enzymes. The major difference between the cobalamin-dependent and -independent algae was found to be associated with methionine metabolism. *C. reinhardtii* and *C. merolae* contain vitamin B₁₂-independent methionine synthase genes (*metE*). In contrast, the vitamin B₁₂-dependent alga *T. pseudonana* contains the gene for the vitamin B₁₂-dependent methionine synthase (*metH*) only. Interestingly, *C. reinhardtii* contains both isoforms of methionine synthase. We used RT-PCR to investigate whether the two *C. reinhardtii* genes were expressed. The results clearly demonstrate that *metH* is expressed both in the presence and in the absence of vitamin B₁₂ (Fig. 1b, tracks marked H), whereas *metE* is expressed only in the absence of the vitamin (track E). It appears that *C. reinhardtii* preferentially uses the vitamin B₁₂-dependent form of methionine synthase, which has a higher rate of catalysis¹⁴, but in the absence of the vitamin the alga is able to survive by inducing the expression of *metE*.

The metabolic consequences of vitamin B₁₂ deprivation in humans is a reduction in methionine synthase activity, causing changes in the balance of intracellular folate derivatives, which ultimately interferes with nucleic acid biosynthesis¹⁵. The biochemical signature of this condition is increased amounts of the substrate homocysteine in blood plasma¹⁶. The relationship between methionine metabolism and vitamin B₁₂ in algae was investigated further using *Lobomonas rostrata*, a vitamin B₁₂-dependent green alga closely related to *C. reinhardtii*. Instead of the vitamin, the alga was grown in the presence of methionine and folate. In the absence of supplements (row I in Fig. 2a), or with methionine (row II) or folate (row III) individually, the cells died after they had been subcultured three

times. However, methionine and folate added together (row IV) rescued vitamin B₁₂ auxotrophy, although the rate of growth was slower than in the presence of vitamin B₁₂ (rows V and VI). Amino acid analysis of *L. rostrata* cells grown in the absence of vitamin B₁₂ showed that homocysteine accumulates in the cell (Fig. 2b), demonstrating that the role of the vitamin in this organism is to make methionine. Thus, the most likely explanation for cobalamin dependency in *L. rostrata* is that it has lost the *metE* gene.

In their natural environment, vitamin B₁₂-dependent algae must be able to obtain the vitamin from an external source. Reported levels of free cobalamin are generally about 2–6 ng l⁻¹ (up to 4 pM) in fresh water¹⁷ and up to 3 ng l⁻¹ (~2 pM) in sea water¹⁸. Our growth studies suggested that this would be insufficient to support algal growth. For example, the marine red alga *Porphyridium purpureum* (Supplementary Fig. S1), the dinoflagellate *Amphidinium operculatum* (Supplementary Fig. S2) and the freshwater euglenoid *Euglena gracilis* (Supplementary Fig. S3) would not grow in minimal media made with natural filter-sterilized sea water or pond water, unless supplemented with at least 10 ng l⁻¹ (~7 pM) vitamin B₁₂. For comparison, most algal culture media contains 10 μ g l⁻¹ vitamin B₁₂. These results demonstrate that the concentration of free cobalamin in the environment is limiting. Given that vitamin B₁₂ dependency has arisen numerous times in the algal kingdom, there must be an alternative, readily accessible source.

Bacteria have previously been implicated in supplying B vitamins to algae (for example, see ref. 17). Indeed, many algal cultures are maintained in culture collections non-axenically. One such culture of *A. operculatum* was obtained from the Culture Collection of Algae and Protozoa (Oban, UK). The algal cells in this culture grew without the addition of exogenous vitamin B₁₂. A species of bacterium was isolated from the culture medium and found to be able to synthesize vitamin B₁₂ *de novo*. The 16S ribosomal RNA gene was amplified by PCR, sequenced and found to be identical to that of *Halomonas* sp. To establish whether this bacterium was responsible for supplying *A. operculatum* with vitamin B₁₂, it was added back to axenic cultures of *A. operculatum* and *P. purpureum*. In both cases, *Halomonas* sp. was able to support growth of the algae to the same extent as vitamin B₁₂ (Supplementary Figs S1 and S2).

Because the medium of these co-cultures did not contain an organic carbon source, the bacteria were presumably using the

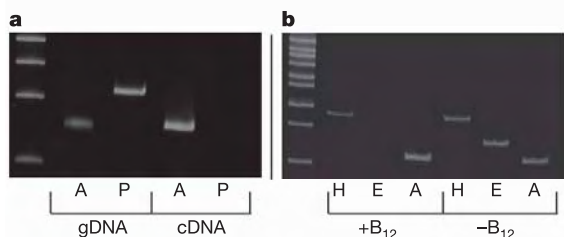


Figure 1 | RT-PCR analysis of genes involved in vitamin B₁₂ metabolism. **a**, Analysis of the gene for cobyrinic acid α , γ -diamide synthase (*cbiP*) from *Thalassiosira pseudonana*. PCR was carried out with either genomic DNA (gDNA) or cDNA as template. Track A used primers to the constitutively expressed actin gene; track P used primers to the *cbiP* gene. The lack of a band in track P with cDNA suggests that the cobyrinic acid α , γ -diamide synthase gene is not expressed. **b**, Analysis of the vitamin B₁₂-dependent (*metH*) and -independent (*metE*) methionine synthase genes from *Chlamydomonas reinhardtii* grown in the presence (+B₁₂, 10 μ g l⁻¹) or absence (-B₁₂) of vitamin B₁₂. Track H used primers for *metH*; Track E used primers for *metE*; Track A used primers for the constitutively expressed actin gene. The lack of a band in track E (+B₁₂) reveals that *metE* is not expressed in the presence of vitamin B₁₂. Details of primers used are given in Supplementary Methods.

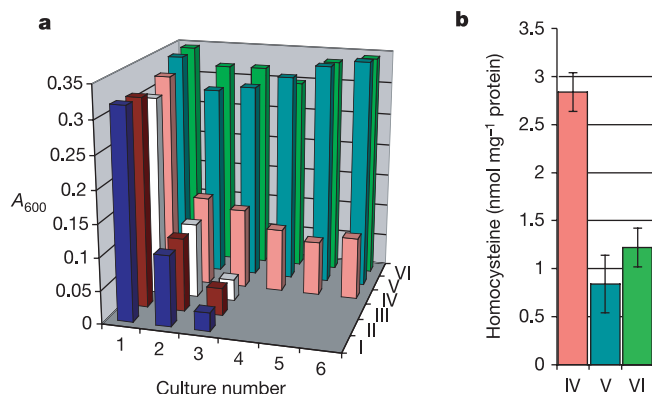


Figure 2 | Effect of vitamin B₁₂ deprivation on the green alga *L. rostrata*. **a**, Growth of *L. rostrata* in Jaworski's medium for six subcultures. I, no supplements; II, 10 mM methionine; III, 1 μ M folate; IV, 1 μ M folate and 10 mM methionine; V, 10 μ g l⁻¹ vitamin B₁₂, 1 μ M folate and 10 mM methionine; VI, 10 μ g l⁻¹ vitamin B₁₂. The values represent the optical density of the cultures measured at 600 nm (A_{600}) after 7 days in continuous light. **b**, Homocysteine analysis of *L. rostrata* cells grown under the indicated conditions in **a** (IV, V and VI). The analysis was repeated three times; error bars represent standard deviation of three replicates. The amount of homocysteine increases in the absence of vitamin B₁₂ because methionine synthase is inactive.

products of algal photosynthesis to grow, suggesting that this is a mutualistic relationship. Examination of cells of *P. purpureum* grown in co-culture with *Halomonas* sp. by microscopy clearly shows the bacterium associated with the extracellular muciferous layer of *P. purpureum* (Fig. 3a). The algal and bacterial cells remain together through multiple washing steps during DAPI staining, indicating that the bacterium is tightly associated with this muciferous layer. A further indication of the intimate relationship was obtained when *Halomonas* sp. was grown in minimal medium in the presence and absence of fucoidin, a commercially available algal extract. The extract increased the rate of bacterial growth, and at the same time significantly increased the amount of vitamin B₁₂ produced from the bacterial cells (Fig. 3b). This suggests that vitamin B₁₂ biosynthesis is upregulated in *Halomonas* sp. in the presence of algal extracts, and that the products of algal metabolism affect the rate of bacterial growth.

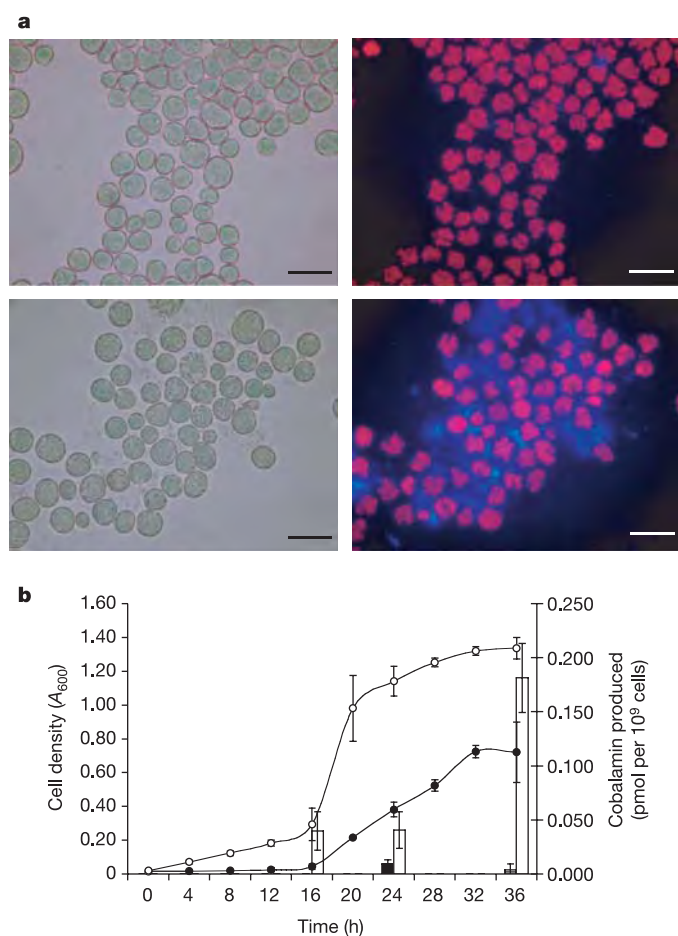


Figure 3 | Interactions between bacteria and algae grown in co-culture. **a**, The association of the red alga *P. purpureum* with *Halomonas* sp. The top left panel is a light microscope image of an axenic *P. purpureum* culture after DAPI staining (see Supplementary Methods). The top right panel is an epifluorescence image of the top left panel. The red chlorophyll epifluorescence masks that produced by DAPI-stained DNA in the algal cells. The bottom left panel is a light microscope image of *P. purpureum* cells grown in co-culture with *Halomonas* sp. after DAPI staining. The bottom right panel is an epifluorescence image of the bottom left panel. In this case there is more blue epifluorescence owing to DAPI staining of bacterial DNA. Scale bars, 30 μ m. **b**, Growth curve with *Halomonas* sp. in ASWH medium grown in the presence (open circles) or absence (filled circles) of fucoidin, a commercially available algal extract. The bars represent the amount of cobalamin produced per 10^9 cells. The presence of fucoidin increases both the rate of bacterial growth and the amount of vitamin B₁₂ produced from the bacterial cells. Error bars represent standard deviation of three replicates.

In conclusion, the data show that over half of all algal species surveyed require vitamin B₁₂ for growth, because, like animals, they require it to make methionine. There is no evolutionary relationship for vitamin B₁₂-dependence in the algae, so it must have arisen numerous times during evolution. There is no evidence of vitamin B₁₂ synthesis in any algal lineage, so vitamin B₁₂ auxotrophy must have resulted from a change in algal metabolism, most likely the loss of the vitamin B₁₂-independent methionine synthase MetE. Data from sequenced algal genomes further support this hypothesis. Most prokaryotes contain both isoforms of methionine synthase, whereas animals have evolved to contain MetH only, and plants MetE only. Therefore, early eukaryotes probably contained both isoforms of methionine synthase, and only later lost one of the genes. Interestingly, the genome of *T. pseudonana* also contains a gene for the vitamin B₁₂-dependent enzyme methylmalonyl CoA mutase, and this enzyme has recently been purified from *Pleurochrysis carterae*¹⁹. Furthermore, *E. gracilis* has been suggested to contain a vitamin B₁₂-dependent type II ribonucleotide reductase²⁰. The presence of these other cobalamin-dependent enzymes may create a further selective pressure to retain vitamin B₁₂ in the metabolism of some species with complex plastids (Table 1).

We have demonstrated that the source of vitamin B₁₂ for microalgae is through a direct interaction with bacteria. We propose that the nature of this interaction is symbiotic, with the algae supplying fixed carbon in return for vitamin B₁₂. Recent studies have shown that macroalgae also have bacteria closely associated with them²¹, suggesting that they too may acquire vitamin B₁₂ by this means. Owing to their colonization of the oceans, algae are responsible for approximately 50% of the world's atmospheric carbon fixation²². Our data indicate that more than half of these are dependent on bacteria for an essential micronutrient. There is also emerging evidence for tantalizingly similar interactions between algae and bacteria for the acquisition of other micronutrients^{23,24}. The recognition of these bitrophic interactions will necessitate a change in our understanding of algal communities, and is likely to have profound implications for the exploitation of algae, both as a food source and for biotechnological applications.

METHODS

Additional methods are given in Supplementary Information.

Growth of algae. The media and conditions used for the growth of different algal strains are provided in the Supplementary Information. With the exception of *Euglena* minimal medium and *Cyanidium* medium²⁵, used to assess the vitamin B₁₂ requirements of *Euglena gracilis* and *Cyanidium caldarium*, respectively, the recipes for all of the other media used in this study can be found on the CCAP website (<http://www.ife.ac.uk/CCAP/>). To assess the growth of different species at natural vitamin B₁₂ concentrations, algal growth media were made exactly as described, except that distilled water was replaced by natural water, and the media filter-sterilized through a 0.2- μ m filter rather than being autoclaved.

Assessment of vitamin B₁₂ requirement of algal species. To assess the vitamin B₁₂ requirement of algal species, 1 ml of culture was used to inoculate 50 ml fresh medium that contained no vitamin B₁₂. The cells were grown for 3–28 days (depending on the species) and then subcultured into fresh medium. This process was repeated five times. If the algal cells died at any point during this process, but remained alive in control medium to which vitamin B₁₂ was added, the species was regarded as requiring vitamin B₁₂. If the cells did not die, the culture was checked to ensure that it was axenic by light microscopy. In addition, the algal culture was plated onto ASWH agar (see Supplementary Information) for marine algae and LB agar for freshwater algae, and checked for bacterial growth. The vitamin B₁₂ content was then measured in cells pelleted from a 50-ml culture. The cell pellet was resuspended in 200 μ l sterile distilled water, heated to 100 $^{\circ}$ C for 10 min, cooled on ice, and then centrifuged at 13,000 g for 2 min. The supernatant was collected and assayed for vitamin B₁₂ by bioassay using *Salmonella typhimurium* strain AR3612 (ref. 26).

Supplementation of *L. rostrata* medium with methionine. *Lobomonas rostrata* was grown in 50 ml Jaworski's medium for 7 days in continuous light with shaking in the presence of 10 μ g l⁻¹ vitamin B₁₂, 10 mM methionine and 1 μ M folic acid. The alga was subcultured six times. After each 7-day period, a 1-ml sample of culture was removed and the optical density measured at 600 nm. For

homocysteine analysis, a 40-ml sample of culture, grown for 7 days, was centrifuged at 5,000 g for 20 min. The supernatant was discarded, and the pellet resuspended in 1 ml distilled water. The cells were sonicated with 6 × 20 s bursts separated by 30 s, on ice, in the presence of 100 µl glass beads (213–300 µm). The cell debris and glass beads were removed by centrifuging the sample at 13,000 g for 5 min at 4 °C, and the supernatant was used for homocysteine analysis (see Supplementary Methods).

Isolation and identification of bacteria. Bacteria were isolated from non-axenic algal cultures by spreading the culture onto ASWH agar plates and incubating them at 30 °C for 36 h. To isolate bacteria from fresh water, a 200-µl sample of the water was spread onto LB agar plates, and the plates were incubated at 20 °C for 48 h. DNA was extracted from bacterial cells, and the 16S rRNA gene was amplified by PCR using degenerate primers as described previously²⁷. The amplified product was gel purified and sequenced directly using the same primers that were used for PCR amplification. The sequence was compared to other sequences in the ribosomal database at Michigan State University (<http://rdp.cme.msu.edu>) to identify the bacterial species.

Generation of axenic *A. operculatum* cultures. *Amphidinium operculatum* cultures were grown in artificial sea water (ASW) at 18 °C on a 16 h/8 h light/dark cycle for 28 days without shaking. A 1-ml aliquot of this culture was used to inoculate ASW containing a cocktail of antibiotics (100 µg ml⁻¹ carbenicillin, 100 µg ml⁻¹ kanamycin and 100 µg ml⁻¹ streptomycin). The *A. operculatum* cells were subcultured three times, for a total of 84 days, in this medium before being transferred back to ASW without antibiotics. To ensure that the final culture was axenic, the algal cells were viewed by light microscopy and plated onto ASWH agar plates. Bacterial contamination was not observed in either case.

Growth experiments with *Halomonas* sp. *Halomonas* sp. was grown aerobically, in ASWH media, at 30 °C in the presence or absence of 0.01% w/v fucoidin (Sigma). The optical density was measured every 4 h. After 16, 24 and 36 h a sample was taken from the culture: 100 µl was used for serial dilutions on ASWH agar plates, and the cells from the remaining 10 ml of medium were assayed for vitamin B₁₂ (as above).

Received 18 May; accepted 15 July 2005.

- Minot, G. R. & Murphy, W. P. Treatment of pernicious anemia by a special diet. 1926. *Yale J. Bio. Med.* **74**, 341–353 (2001).
- Seetharam, B., Bose, S. & Li, N. Cellular import of cobalamin (Vitamin B₁₂). *J. Nutr.* **129**, 1761–1764 (1999).
- Seetharam, B. & Alpers, D. H. Absorption and transport of cobalamin (vitamin B₁₂). *Annu. Rev. Nutr.* **2**, 343–369 (1982).
- Banerjee, R. & Ragsdale, S. W. The many faces of vitamin B₁₂: catalysis by cobalamin-dependent enzymes. *Annu. Rev. Biochem.* **72**, 209–247 (2003).
- Watanabe, F. *et al.* Characterization of a vitamin B₁₂ compound in the edible purple laver, *Porphyra yezoensis*. *Biosci. Biotech. Biochem.* **64**, 2712–2715 (2000).
- Stabler, S. P. & Allen, R. H. Vitamin B₁₂ deficiency as a worldwide problem. *Annu. Rev. Nutr.* **24**, 299–326 (2004).
- Kanazawa, A. Vitamins in algae. *Bull. Jap. Soc. Sci. Fish.* **29**, 713–731 (1963).
- Warren, M. J., Raux, E., Schubert, H. L. & Escalante-Semerena, J. C. The biosynthesis of adenosylcobalamin (Vitamin B₁₂). *Nat. Prod. Rep.* **19**, 390–412 (2002).
- Carlucci, A. F. & Bowes, P. M. Vitamin production and utilization by phytoplankton in mixed culture. *J. Phycol.* **6**, 393–400 (1970).
- Provasoli, L. & Carlucci, A. F. in *Algal Physiology and Biochemistry* (ed. Stewart, W. D. P.) 741 (Blackwell, Oxford, 1974).
- Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657 (2004).
- Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
- Watanabe, F., Nakano, Y., Tamura, Y. & Yamanaka, H. Vitamin B₁₂ metabolism in a photosynthesizing green alga, *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta* **1075**, 36–41 (1991).
- Gonzalez, J. C., Banerjee, R. V., Huang, S., Summer, J. S. & Matthews, R. G. Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: two solutions to the same chemical problem. *Biochemistry* **31**, 6045–6056 (1992).
- Scott, J. & Weir, D. Folate/vitamin B₁₂ inter-relationships. *Essays Biochem.* **28**, 63–72 (1994).
- Cantu, C. *et al.* Hyperhomocysteinemia, low folate and vitamin B₁₂ concentrations, and methylene tetrahydrofolate reductase mutation in cerebral venous thrombosis. *Stroke* **35**, 1790–1794 (2004).
- Kurata, A. in *Chrysophytes: Aspects and Problems* (eds Kristiansen, J. & Andersen, R. A.) 185–196 (Cambridge Univ. Press, Cambridge, 1986).
- Carlucci, A. F. The ecology of the plankton off La Jolla, California in the period April through September, 1967. II. Vitamin B₁₂, thiamine and biotin. *Bull. Sci. Inst. Ocean* **17**, 23–30 (1970).
- Miyamoto, E., Watanabe, F., Yamaguchi, Y., Takenaka, H. & Nakano, Y. Purification and characterization of methylmalonyl-CoA mutase from a photosynthetic coccolithophorid alga, *Pleurochrysis carterae*. *Comp. Biochem. Physiol. B* **138**, 163–167 (2004).
- Hamilton, F. D. Ribonucleotide reductase from *Euglena gracilis*. A 5'-deoxyadenosylcobalamin-dependent enzyme. *J. Biol. Chem.* **249**, 4428–4434 (1974).
- Murakami, A., Miyashita, H., Iseki, M., Adachi, K. & Mimuro, M. Chlorophyll *d* in an epiphytic cyanobacterium of red algae. *Science* **303**, 1633 (2004).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Butler, A. Acquisition and utilization of transition metal ions by marine organisms. *Science* **281**, 207–210 (1998).
- Keshtacher-Liebson, E., Hadar, J. T. & Chen, Y. Oligotrophic bacteria enhance algal growth under iron-deficient conditions. *Appl. Environ. Microbiol.* **61**, 2439–2441 (1995).
- Troxler, R. F. Synthesis of bile pigments in plants. Formation of carbon monoxide and phycocyanobilin in wild-type and mutant strains of the alga, *Cyanidium caldarium*. *Biochemistry* **11**, 4235–4242 (1972).
- Raux, E. *et al.* *Salmonella typhimurium* cobalamin (vitamin B₁₂) biosynthetic genes: functional studies in *S. typhimurium* and *Escherichia coli*. *J. Bacteriol.* **178**, 753–767 (1996).
- Lane, D. J. *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115 (John Wiley and Sons, New York, 1991).
- Durnford, D. G. *et al.* A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J. Mol. Evol.* **48**, 59–68 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Howe, S. Purton, S. Beale and D. Vernon for the donation of algal strains, the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK for the award of an earmarked studentship to M.T.C., the European Union Viteomics Research Training Network for funding and for providing a forum for discussions, and Queen Mary University of London for providing a studentship to A.D.L.

Author Contributions This work is the result of a collaboration between the laboratories of M.J.W. and A.G.S. through the joint earmarked studentship for M.T.C.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.G.S. (as25@cam.ac.uk).

LETTERS

The transcription factor Engrailed-2 guides retinal axons

Isabelle Brunet^{1,2}, Christine Weinl², Michael Piper², Alain Trembleau¹, Michel Volovitch¹, William Harris², Alain Prochiantz¹ & Christine Holt²

Engrailed-2 (En-2), a homeodomain transcription factor, is expressed in a caudal-to-rostral gradient in the developing mid-brain, where it has an instructive role in patterning the optic tectum—the target of topographic retinal input^{1,2}. In addition to its well-known role in regulating gene expression through its DNA-binding domain, En-2 may also have a role in cell–cell communication, as suggested by the presence of other domains involved in nuclear export, secretion and internalization³. Consistent with this possibility, here we report that an external gradient of En-2 protein strongly repels growth cones of *Xenopus* axons originating from the temporal retina and, conversely, attracts nasal axons. Fluorescently tagged En-2 accumulates inside growth cones within minutes of exposure, and a mutant form of the protein that cannot enter cells fails to elicit axon turning. Once internalized, En-2 stimulates the rapid phosphorylation of proteins involved in translation initiation and triggers the local synthesis of new proteins. Furthermore, the turning responses of both nasal and temporal growth cones in the presence of En-2 are blocked by inhibitors of protein synthesis. The differential guidance of nasal and temporal axons reported here suggests that En-2 may participate directly in topographic map formation in the vertebrate visual system.

The first molecular insight into the mechanism of topographic mapping in the nervous system is usually credited to Sperry⁴, whose chemoaffinity hypothesis proposed matching gradients of receptors and ligands within the retina and tectum. The first candidate molecules fulfilling the requirements of this hypothesis, the EphrinA ligands, were identified in the tectum and found to be repulsive to retinal axons expressing EphA receptors^{5,6}. Temporal axons expressing high levels of EphA receptors map to the rostral tectum and avoid the EphrinA-rich caudal tectum. En-2, which is also expressed in a caudal-to-rostral gradient in the developing tectum, has been shown to promote the expression of tectal EphrinA ligands^{7,8} and, through its transcriptional activity, is thought to have a major role in setting up the EphrinA gradient. However, work in knockout mice has shown that in the total absence of EphrinAs, a rough retinotectal map still forms^{9,10}, implicating the existence of other potential guidance cues. In light of the evidence that En-2 can be secreted and transferred from one cell to another³, we decided to test whether En-2 could also act as a guidance factor.

We first asked whether a source of pure En-2 protein could guide *Xenopus* retinal axons. Retinal growth cones from nasal and temporal parts of the retina were individually exposed to gradients of En-2 in chemotropic turning assays (see Methods)^{11,12}. Notably, nasal axons were attracted by En-2, whereas temporal axons were repelled (Fig. 1a–e). These opposite turning responses correspond to the *in vivo* organization of the retinotectal map, where nasal axons terminate in the En-2-rich caudal tectum but temporal axons

avoid it. Similar results were obtained when the experiment was performed ‘blind’ with respect to the nasal or temporal origin of the axons (Supplementary Fig. 1). The effect was not due to differential growth rates in the presence of En-2, as both nasal and temporal

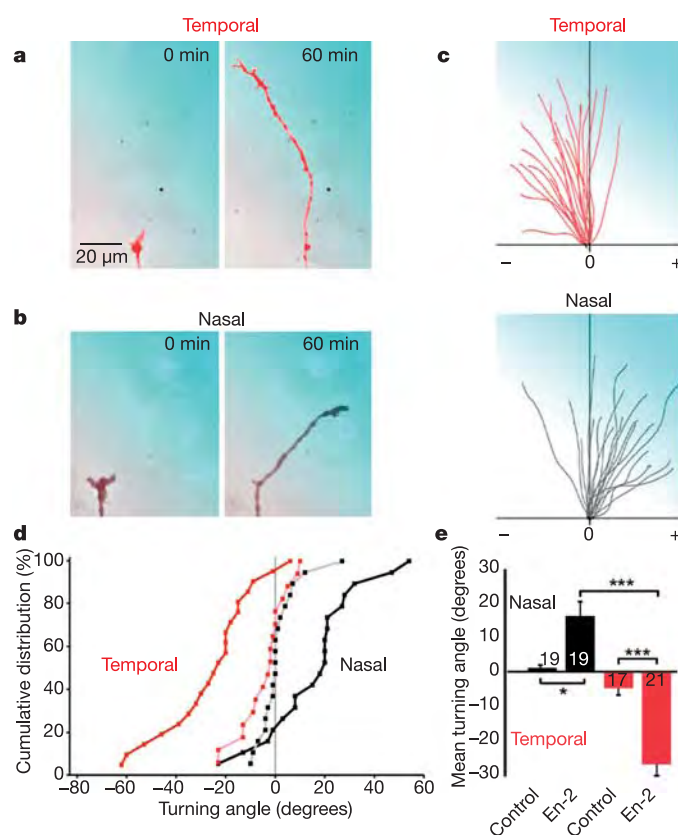


Figure 1 | En-2 gradient repels temporal and attracts nasal retinal axons. **a, b**, Examples of temporal (red) and nasal (black) growth cones tested in turning assays with $10 \mu\text{g ml}^{-1}$ En-2 in the pipette (300 pM at the growth cone). The pipette is positioned in the top right, and the En-2 gradient is represented in blue (**a–c**). **c**, Trajectory plots of temporal (red) and nasal (black) neurites in En-2 gradients. Each line represents a single growth cone trajectory; the origin represents the centre of the growth cone at 0 min, and positive (+) and negative (–) turning angles are indicated. **d**, Cumulative distributions of turning angles of temporal (red) and nasal (black) growth cones in En-2 (bold) and control (light) gradients. **e**, Mean turning angles of growth cones in the experimental conditions shown in **d**, numbers on or beside the bars denote the number of growth cones tested. Significance was calculated using a Kolmogorov–Smirnov test, and indicated by asterisks (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). Error bars indicate s.e.m.

¹CNRS UMR 8542, Ecole Normale Supérieure, 46 rue d’Ulm, 75230 Paris Cedex 05, France. ²Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.

axons grew at approximately $30 \mu\text{m h}^{-1}$ in these assays (Supplementary Fig. 2). These data demonstrate that En-2 can act as a soluble factor that differentially guides axons from temporal and nasal parts of the retina.

As the En-2 protein possesses an internalization sequence¹³, we asked whether En-2 was acting within the retinal growth cones to direct their growth. To visualize internalization, we exposed growth cones to media containing fluorescein isothiocyanate (FITC)-tagged En-2 for 5 min and then washed them in media free of FITC-En-2. Within 15 min, a fluorescent signal accumulated in living growth cones, occasionally residing in filopodia and lamellipodia, but mostly concentrated in puncta in the body of the growth cone and along the neurite shaft (Fig. 2a). Previous studies have shown that internalization of En-2 is compromised when there is a mutation in the coding sequence of the penetratin domain, changing the tryptophan (W) and phenylalanine (F) residues at positions 48 and 49 in the homeodomain to serine (S) and arginine (R) residues, respectively

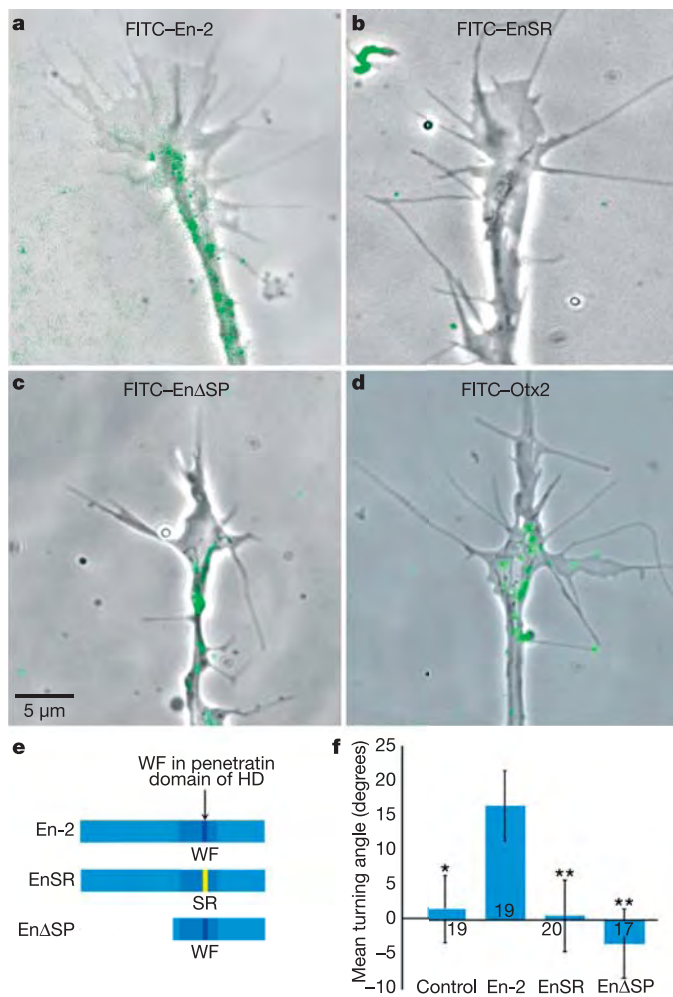


Figure 2 | En-2 is internalized in the growth cone and guidance depends on internalization. **a–c**, Live retinal growth cones following 5 min exposure to FITC-tagged proteins. Over 50% of growth cones showed internalized fluorescent puncta (green) following exposure to FITC-En-2 (**a**) or FITC-EnΔSP (**c**), but not with FITC-EnSR (**b**). **d**, FITC-Otx2 is internalized. **e**, En-2 constructs used in the turning assays: En-2 is the full-length wild-type protein; EnSR has a mutation in the penetratin domain (amino acids WF replaced by SR); and EnΔSP lacks the N-terminal part of the protein (where a putative eIF4E-binding site resides). HD, homeodomain. **f**, Mean turning angles of nasal growth cones in gradients of En-2, EnSR and EnΔSP. Nasal axons turn towards En-2 but not EnSR or EnΔSP. For Otx2 activity, see Supplementary Fig. 1d. For *P* values, see Fig. 1 legend, and comparisons are to En-2. Error bars indicate s.e.m.

(W48S-F49R; see Fig. 2e)¹⁴. A FITC-tagged version of this mutant protein (FITC-EnSR) did not enter growth cones, even at high doses (Fig. 2b). Therefore, we used the EnSR mutant protein to test whether En-2 needs to be internalized to elicit axon turning. When nasal retinal growth cones were exposed to a gradient of the EnSR mutant protein they did not exhibit any directional turning bias (Fig. 2f), strongly suggesting that internalization is critical for axon guidance.

Turning behaviour was usually initiated in response to En-2 within 15–20 min of exposure—too rapidly for a transcriptional role of En-2 in axon guidance, especially as the growth cone is often $>100 \mu\text{m}$ from the soma, or cell body. Recent studies have shown that the homeoprotein HoxA9, and a large number of other homeodomain proteins, contain a highly conserved binding site to eukaryotic initiation factor 4E (eIF4E) that is typically found in translational regulators^{15,16}. Of particular note, HoxA9 has been found to positively regulate translation by directly competing with factors that bind and repress eIF4E (ref. 16). Extracellular guidance cues, such as Semaphorin 3A and Netrin-1, lead to local protein synthesis within retinal growth cones, and this rapidly activated translation is essential for chemotropic turning in gradients of these cues¹⁷. Therefore, we examined the possibility that En-2 stimulates turning responses in retinal growth cones by affecting local translation. First, a set of experiments was conducted to rule out a transcriptional mechanism of En-2 action. Nasal growth cones transected from their cell bodies still turned towards an En-2 gradient, and isolated temporal growth cones continued to turn away from this gradient (Fig. 3a and Supplementary Fig. 3). We then asked whether these transected and isolated growth cones synthesized new proteins in response to En-2 by measuring the incorporation of tritiated (³H)-leucine. Indeed, 10 min of stimulation with En-2 caused isolated growth cones to incorporate significantly higher amounts of ³H-leucine than controls, and this incorporation was severely reduced by the protein synthesis inhibitor anisomycin, but not the transcription inhibitor α -amanitin (Fig. 3b). En-2 binds eIF4E (ref. 18), but the EnΔSP mutant¹⁹ lacks the amino-terminal region of the protein (Fig. 2e) that contains a putative eIF4E-binding domain. The FITC-tagged EnΔSP protein is internalized normally (Fig. 2c), but in turning assays, gradients of EnΔSP do not attract nasal axons (Fig. 2f), consistent with the hypothesis that En-2-induced turning requires the N-terminal domain. Finally, another set of experiments was conducted to test whether pharmacological reagents that specifically interfere with transcription or translation affected growth cone turning in response to En-2. As shown in Fig. 3c, anisomycin abolished the ability of nasal axons to turn towards, and temporal axons to turn away, from a source of En-2. Similarly, rapamycin—an antibiotic that blocks the translation of 5'-capped messenger RNA by inhibiting target of rapamycin (TOR)—strongly inhibited axon turning responses. In contrast, α -amanitin—an inhibitor of transcription—had no effect on En-2-induced axon turning. Together, these results suggest that translation, rather than transcription, is essential for En-2-mediated retinal axon guidance in *Xenopus*.

TOR-dependent translation can be regulated by eIF4E-binding protein 1 (4E-BP1), which silences translation by competitively binding eIF4E and sequestering it from the translation initiation complex²⁰. Phosphorylation of 4E-BP1 by TOR releases eIF4E and allows it to initiate translation²¹. In growth cones exposed to Semaphorin 3A or Netrin-1, a rapid rise in the phosphorylation of both eIF4E and 4E-BP1 has been reported using antibodies that recognize the phosphorylated forms of the proteins¹⁷. Therefore, we postulated that if En-2 was acting through this pathway, then 4E-BP1 should also be phosphorylated in growth cones exposed to En-2. We compared the levels of phospho-4E-BP1 and phospho-eIF4E in control and En-2-exposed retinal growth cones using phospho-specific antibodies (Fig. 4). Digital quantification revealed that En-2 triggers a significant rise in both phospho-4E-BP1 and phospho-eIF4E immunofluorescence levels after 5 min in temporal and

nasal growth cones (Fig. 4a, b). In contrast, En Δ SP—even though it is internalized (Fig. 2c)—does not stimulate the phosphorylation of either of these translation-associated proteins (Fig. 4c). These data are consistent with the possibility that En-2 activates translation in growth cones by interacting with components of the translational pathway.

A key question is that of specificity. Do all homeodomain proteins elicit the same response, or are these responses specific to En-2? To address this, we tested Otx2, another homeodomain transcription factor that shares the secretion, internalization and eIF4E-binding properties of En-2 (ref. 18). *In vivo*, Otx2 is distributed fairly uniformly in the developing *Xenopus* forebrain and midbrain²². FITC–Otx2 is internalized in retinal growth cones (Fig. 2d). However, when applied in a gradient, Otx2 weakly attracted both the nasal and temporal populations of retinal growth cones (Supplementary Fig. 4). Thus, unlike En-2, Otx2 does not elicit opposite turning

responses in the two populations of retinal axons, indicating that En-2 acts specifically to induce these divergent behaviours.

Not only is the caudal-to-rostral gradient of En-2 in the tectum perfectly suited to having a function in topographic mapping, but the non-transcriptional function of En-2 can explain previously unresolved issues in retinotectal mapping, such as what attracts nasal axons to the caudal tectum or to patches of the tectum that retrovirally overexpress En-2 (ref. 23). Our findings may also help to explain why a fair degree of order remains along the rostrocaudal axis in the retinotectal maps of EphrinA2/EphrinA5 double knockout mice⁹, although they do not challenge the view that EphrinA ligands have a major role in retinotopic guidance. Rather, they suggest that En-2 may complement this role by eliciting different responses in nasal versus temporal axons. It is not clear how En-2 exerts its differential effect. One possibility is that En-2 triggers the translation of different proteins in the nasal and temporal axon populations.

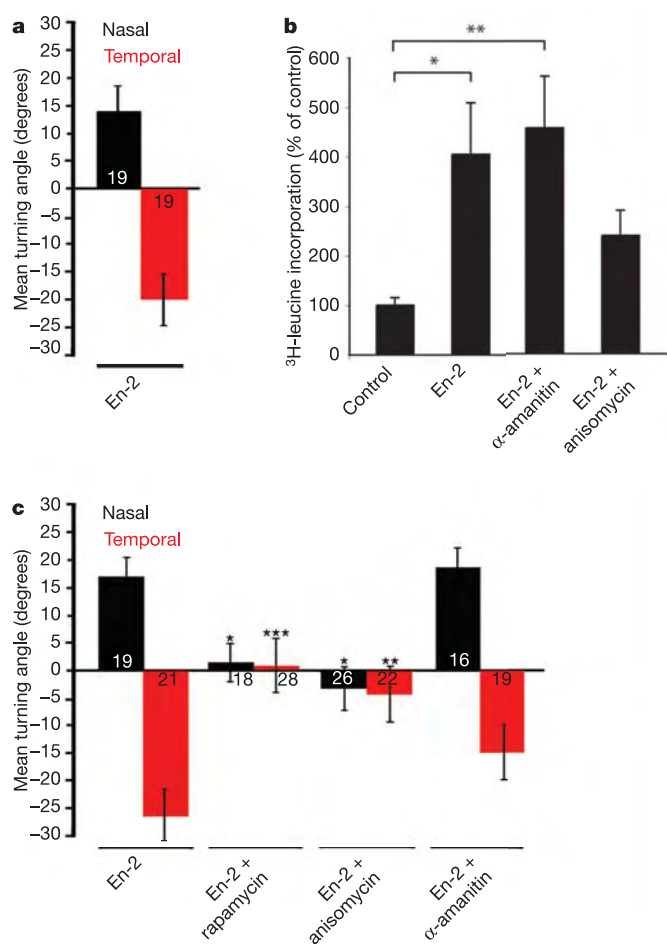


Figure 3 | Retinal ganglion cell growth cone guidance by En-2 is dependent on protein synthesis. **a**, Temporal (red) and nasal (black) growth cones isolated from their somas (see Supplementary Fig. 1c) show opposite turning responses to an En-2 gradient, eliminating a role for transcription. **b**, ³H-leucine incorporation in growth cones isolated from their cell bodies. En-2 stimulates significant incorporation of ³H-leucine, which is abolished by anisomycin but not α -amanitin. **c**, Mean turning angles of temporal and nasal growth cones in a gradient of En-2 in the presence of inhibitors. En-2-induced turning is blocked by anisomycin and rapamycin, but is unaffected by α -amanitin. The number of growth cones tested in each case is indicated on the bars. **b**, Significance was calculated using a Kruskal–Wallis test, and indicated by asterisks (* P < 0.05; ** P < 0.01). **c**, Significance was calculated using a Kolmogorov–Smirnov test, and indicated by asterisks (* P < 0.05; ** P < 0.01; *** P < 0.001), and comparisons are to En-2. Error bars indicate s.e.m.

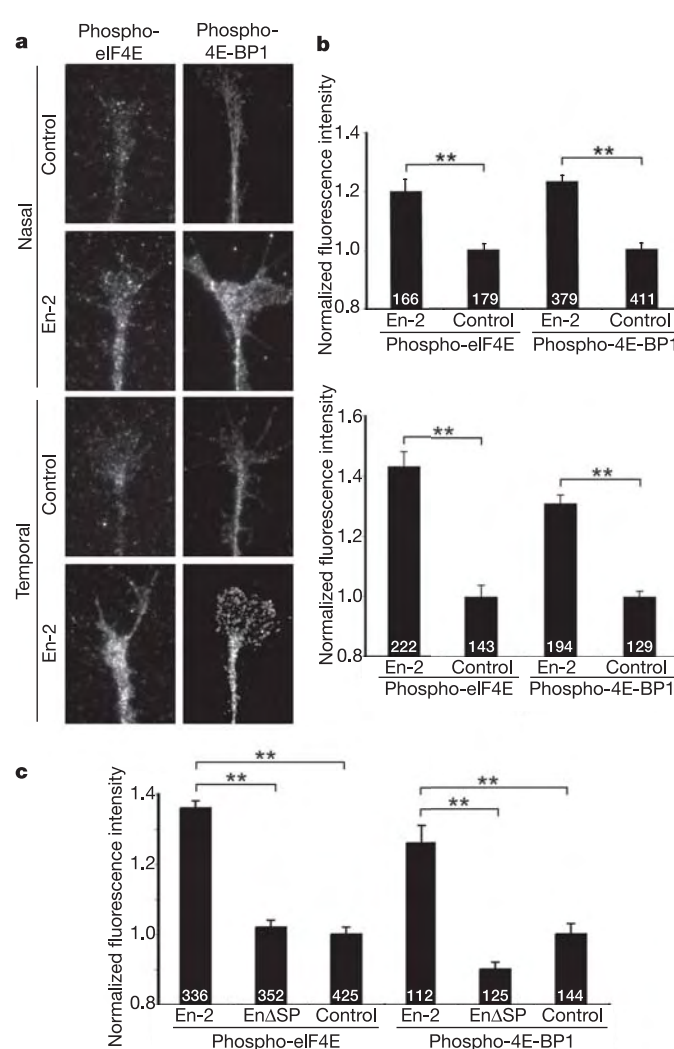


Figure 4 | En-2 stimulates the phosphorylation of translational regulatory proteins. **a**, Fluorescent micrographs of unstimulated control and En-2-stimulated growth cones using phospho-eIF4E and phospho-eIF4E-BP1 antibodies. Exposure to En-2 for 5 min increases the level of phosphorylation of eIF4E and eIF4E-BP1 in nasal and temporal growth cones. **b**, Histograms showing the digital quantification of the fluorescence signal in growth cones. **c**, Exposure of growth cones to En Δ SP does not lead to phosphorylation of either eIF4E or eIF4E-BP1. The number of growth cones tested in each case is indicated on the bars. Significance was calculated using a two-tailed Mann–Whitney test, and indicated by asterisks (** P < 0.0001). Error bars indicate s.e.m.

Alternatively, En-2 may activate translation of the same proteins, but in a spatially opposite manner. Providing answers to these questions will require further investigation.

Recent reports have demonstrated a role for morphogens like Wnt, Bone morphogenetic protein and Hedgehog in axonal guidance²⁴, and so perhaps it is not unexpected that if a homeodomain protein is secreted, and its extracellular concentration reflects its pattern of expression in the brain, then such concentration gradients could be also used to guide growing axons. Notably, previous studies have shown that there is a high degree of correlation between the expression domains of homeodomain transcription factors and early axonal pathways²⁵. In the context of the results presented here, the possibility that many of these richly patterned transcription factors have a direct role in axonal guidance should not be overlooked.

METHODS

Xenopus embryos. Embryos were obtained by *in vitro* fertilization of oocytes from *Xenopus laevis* females stimulated with human chorionic gonadotropin (Sigma). Embryos were raised in 0.1 × Modified Barth's Saline, pH 7.4, and were staged according to the tables of Nieuwkoop and Faber.

Growth cone turning assays. The extreme nasal and temporal thirds of stage 32–33 retinas were explanted onto 50-mm glass-bottomed Petri dishes (MaTek) precoated with 10 µg ml⁻¹ poly-L-lysine followed by 10 µg ml⁻¹ Laminin-1 (Sigma), as this substrate promotes good outgrowth from retinal explants and is thought to be a substrate for retinal axon growth *in vivo*²⁶. Cultures were grown at 20°C for 12–26 h to generate good neurite outgrowth¹⁷. Stable gradients of En-2 protein were produced as described previously^{11,12} by pulsatile ejection of En-2 protein (10 µg ml⁻¹). The pipette tip was positioned 100 µm from the centre of the growth cone at a 45° angle to the direction of growth. These standard conditions give rise to an estimated 1,000-fold dilution of the factor between the pipette and the growth cone, and a concentration gradient across the growth cone of 5–10% (ref. 11). The turning angle for each growth cone was defined as the angle between the original and final direction of growth after 60 min of culture. Only growth cones with a net extension >5 µm over the assay period were included in the analysis.

Pharmacological reagents. The following pharmacological reagents were bath-applied to cultures immediately before the application of En-2 protein in the turning assay: 40 µM anisomycin (Sigma), which inhibits the peptidyl transferase activity on the ribosome; 10 µg ml⁻¹ α-amanitin (Calbiochem), which inhibits RNA polymerase II; and 10 nM rapamycin (Calbiochem), which inhibits mammalian TOR.

Protein production. The En-2, EnSR and EnΔSP proteins were produced in isopropyl-β-D-thiogalactoside (IPTG)-inducible bacteria using published methods²⁷, and Otx2 was produced in insect cells from a recombinant baculovirus (a gift from G. Corte). Protein purity was analysed by SDS-polyacrylamide gel electrophoresis. After concentration measurement, the proteins were stored at -20°C in culture medium supplemented with 10% glycerol. Before use, the glycerol was removed by dialysis against 1 × PBS. FITC was conjugated to purified En2 using standard procedures (FITC-Celite, Sigma).

Measurement of ³H-leucine incorporation. ³H-leucine incorporation in isolated growth cones was assayed following 10 min of stimulation with En-2 according to methods described previously¹⁷. En-2 (0.3 µg ml⁻¹) was added globally to the cultures and the experiments were done in six independent trials, each in duplicate.

FITC-protein internalization. FITC-tagged proteins (FITC-En-2, 2.3 µg ml⁻¹; FITC-EnSR, 20.7 µg ml⁻¹; FITC-EnΔSP, 6.5 µg ml⁻¹; FITC-Otx2, 8.3 µg ml⁻¹) were added to retinal cultures for 5 min after blocking for 30 min with 4% BSA. The cultures were then washed extensively in FITC-protein-free medium, and imaged. Fluorescence was visualized in living growth cones with a ×100 objective on a Nikon Eclipse 2000U microscope, and images were captured with a Hamamatsu digital CCD camera using Openlab software (Improvision). Approximately 30 cultures were examined for each condition, with approximately 40 growth cones per culture. More than half of the growth cones were labelled when exposed to FITC-En-2 or FITC-EnΔSP, but not a single growth cone was found labelled with FITC-EnSR.

Quantitative immunocytochemistry. Antibodies were used at the following dilutions: phospho-4E-BP1 primary antibody (Cell Signaling Technology) at 1:100; phospho-eIF4E primary antibody (Cell Signaling Technology) at 1:25; and goat anti-rabbit Cy3 secondary antibody (Chemicon) at 1:700. *Xenopus* temporal or nasal retinal explants were dissected at stage 32 and grown overnight at 20°C on coverslips coated with poly-L-lysine (10 µg ml⁻¹) and Laminin-1

(10 µg ml⁻¹). Growth cones were stimulated with En-2 (1 µg ml⁻¹) added globally to the medium for 5 min, fixed in 4% paraformaldehyde and 15% sucrose for 30 min, permeabilized in 0.1% Triton-X100, blocked with 10% goat serum, and incubated sequentially in primary and secondary antibodies for 1 h. Coverslips were mounted in FluorSave (Calbiochem) and growth cones were imaged at ×100 on a Nikon Optiphot. For digital quantification, growth cones were selected at random under phase optics and a phase image was captured, followed by a fluorescent image, using a Hamamatsu digital CCD camera and taking care to avoid pixel saturation. The growth cone outline was traced on the phase image, superimposed on the fluorescent image, and the fluorescence intensities were measured digitally using Openlab software yielding values of pixel intensity per unit area. The background fluorescence was measured by placing the outline of the growth cone in an adjacent area devoid of cellular material, and subtracted from the growth cone values to give a background-corrected intensity value. For presentation of the data, the fluorescence intensity values were normalized to the respective control experiments. Each experiment was performed four independent times and run in duplicate.

Statistical analyses. All data are shown as mean ± s.e.m. Significance was assessed using the two-sample Kolmogorov–Smirnov test, unless otherwise stated. Statistical analyses were carried out using the statistical package 'R' (www.r-project.org).

Received 1 June; accepted 4 August 2005.

1. Retaux, S. & Harris, W. A. Engrailed and retinotectal topography. *Trends Neurosci.* **19**, 542–546 (1996).
2. Itasaki, N. & Nakamura, H. A role for gradient expression in positional specification on the optic tectum. *Neuron* **16**, 55–62 (1996).
3. Prochiantz, A. & Joliot, A. Can transcription factors function as cell–cell signalling molecules? *Nature Rev. Mol. Cell Biol.* **4**, 814–819 (2003).
4. Sperry, R. W. Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl Acad. Sci. USA* **50**, 703–710 (1963).
5. Cheng, H. J., Nakamoto, M., Bergemann, A. D. & Flanagan, J. G. Complementary gradients in expression and binding of ELF-1 and Mek4 in development of the topographic retinotectal projection map. *Cell* **82**, 371–381 (1995).
6. Drescher, U. *et al.* *In vitro* guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands for Eph receptor tyrosine kinases. *Cell* **82**, 359–370 (1995).
7. Shigetani, Y., Funahashi, J. I. & Nakamura, H. En-2 regulates the expression of the ligands for Eph type tyrosine kinases in chick embryonic tectum. *Neurosci. Res.* **27**, 211–217 (1997).
8. Logan, C. *et al.* Rostral optic tectum acquires caudal characteristics following ectopic engrailed expression. *Curr. Biol.* **6**, 1006–1014 (1996).
9. Feldheim, D. A. *et al.* Genetic analysis of ephrin-A2 and ephrin-A5 shows their requirement in multiple aspects of retinocollicular mapping. *Neuron* **25**, 563–574 (2000).
10. Frisen, J., Holmberg, J. & Barbacid, M. Ephrins and their Eph receptors: multitasked directors of embryonic development. *EMBO J.* **18**, 5159–5165 (1999).
11. Lohof, A. M., Quillan, M., Dan, Y. & Poo, M. M. Asymmetric modulation of cytosolic cAMP activity induces growth cone turning. *J. Neurosci.* **12**, 1253–1261 (1992).
12. de la Torre, J. R. *et al.* Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor DCC. *Neuron* **19**, 1211–1224 (1997).
13. Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **269**, 10444–10450 (1994).
14. Joliot, A. *et al.* Identification of a signal sequence necessary for the unconventional secretion of Engrailed homeoprotein. *Curr. Biol.* **8**, 856–863 (1998).
15. Topisirovic, I. *et al.* The proline-rich homeodomain protein, PRH, is a tissue-specific inhibitor of eIF4E-dependent cyclin D1 mRNA transport and growth. *EMBO J.* **22**, 689–703 (2003).
16. Topisirovic, I. *et al.* Eukaryotic translation initiation factor 4E activity is modulated by HOXA9 at multiple levels. *Mol. Cell. Biol.* **25**, 1100–1112 (2005).
17. Campbell, D. S. & Holt, C. E. Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation. *Neuron* **32**, 1013–1026 (2001).
18. Nedelec, S. *et al.* Emx2 homeodomain transcription factor interacts with eukaryotic translation initiation factor 4E (eIF4E) in the axons of olfactory sensory neurons. *Proc. Natl Acad. Sci. USA* **101**, 10815–10820 (2004).
19. Foucher, I., Montesinos, M. L., Volovitch, M., Prochiantz, A. & Trembleau, A. Joint regulation of the MAP1B promoter by HNF3β/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. *Development* **130**, 1867–1876 (2003).
20. Hay, N. & Sonenberg, N. Upstream and downstream of mTOR. *Genes Dev.* **18**, 1926–1945 (2004).

21. Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).
22. Kablar, B. *et al.* *Xotx* genes in the developing brain of *Xenopus laevis*. *Mech. Dev.* **55**, 145–158 (1996).
23. Friedman, G. C. & O'Leary, D. D. Retroviral misexpression of *engrailed* genes in the chick optic tectum perturbs the topographic targeting of retinal axons. *J. Neurosci.* **16**, 5498–5509 (1996).
24. Charron, F. & Tessier-Lavigne, M. Novel brain wiring functions for classical morphogens: a role as graded positional cues in axon guidance. *Development* **132**, 2251–2262 (2005).
25. Macdonald, R. *et al.* Regulatory gene expression boundaries demarcate sites of neuronal differentiation in the embryonic zebrafish forebrain. *Neuron* **13**, 1039–1053 (1994).
26. Cohen, J., Burne, J. F., McKinlay, C. & Winter, J. The role of laminin and the laminin/fibronectin receptor complex in the outgrowth of retinal ganglion cell axons. *Dev. Biol.* **122**, 407–418 (1987).
27. Montesinos, M. L. *et al.* The neuronal microtubule-associated protein 1B is under homeoprotein transcriptional control. *J. Neurosci.* **21**, 3350–3359 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all of our colleagues for discussions, and A. Dwivedy and E. Ipendey for technical assistance. This work was supported by the Wellcome Trust, UK (to C.H. and W.H.), the Human Frontier Sciences Program and the European Commission (to A.P.).

Author Contributions I.B. did the turning assays and internalization experiments in Figs 1, 2, 3a and 3c, and in the Supplementary Figs. C.W. did the quantitative immunofluorescence work in Fig. 4, and trained I.B. in turning assays. M.P. did the ³H-leucine experiments in Fig. 3b. M.V. provided all the constructs and proteins. A.P. proposed the hypothesis that En-2 guides retinal axons. A.P. and A.T. initiated the collaboration and discussed experiments. W.H. and C.H. wrote the manuscript and discussed the experiments.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.H. (ceh@mole.bio.cam.ac.uk) or A.P. (prochian@biologie.ens.fr).

Protection of macaques from vaginal SHIV challenge by vaginally delivered inhibitors of virus–cell fusion

Ronald S. Veazey¹, Per Johan Klasse², Susan M. Schader², Qinxue Hu³, Thomas J. Ketas², Min Lu⁴, Preston A. Marx¹, Jason Dufour¹, Richard J. Colonno⁵, Robin J. Shattock³, Martin S. Springer⁶ & John P. Moore²

Human immunodeficiency virus type 1 (HIV-1) continues to spread, principally by heterosexual sex, but no vaccine is available¹. Hence, alternative prevention methods are needed to supplement educational and behavioural-modification programmes. One such approach is a vaginal microbicide: the application of inhibitory compounds before intercourse². Here, we have evaluated the microbicide concept using the rhesus macaque 'high dose' vaginal transmission model with a CCR5-receptor-using simian–human immunodeficiency virus (SHIV-162P3) and three compounds that inhibit different stages of the virus–cell attachment and entry process. These compounds are BMS-378806, a small molecule that binds the viral gp120 glycoprotein and prevents its attachment to the CD4 and CCR5 receptors^{3,4}, CMPD167, a small molecule that binds to CCR5 to inhibit gp120 association⁵, and C52L, a bacterially expressed peptide inhibitor of gp41-mediated fusion⁶. *In vitro*, all three compounds inhibit infection of T cells and cervical tissue explants, and C52L acts synergistically with CMPD167 or BMS-378806 to inhibit infection of cell lines. *In vivo*, significant protection was achieved using each compound alone and in combinations. CMPD167 and BMS-378806 were protective even when applied 6 h before challenge.

The compounds BMS-378806, CMPD167 and C52L potently inhibit SHIV-162P3 infection of macaque and human peripheral blood mononuclear cells (PBMCs) *in vitro*, with half-maximal effective concentration (EC₅₀) values in the low nanomolar range, and they efficiently block infection of cervical tissue explants and monocyte-derived macrophages (MDMs) by the R5 isolate HIV-1_{BaL} (Table 1). Their inhibitory activity in combination is reported as Supplementary Information.

A microbicide must be safe. Several small-molecule CCR5 inhibitors are now in phase II or III trials for treatment of HIV-1 infection, without significant safety concerns yet arising⁷. We have administered CMPD167 systemically to macaques, again without problems^{8,9}. BMS-378806 and a subsequent compound, BMS-388043, were successfully tested for safety in humans, and more potent, related compounds are undergoing trials¹⁰. C52L is a sequence-modified version of the approved drug Enfuvirtide¹¹. No local irritation or inflammation was observed upon colposcopy and biopsy examination of the macaque vagina after multiple applications of CMPD167, BMS-378806 or C52L, and, at concentrations inhibitory to HIV-1 infection *in vitro*, they had no adverse effects on human cervical explant viability (data not shown).

The SHIV-162P3 challenge virus resembles most naturally

transmitted HIV-1 strains by using CCR5 (refs 12, 13). The infection rate for our stock in control animals was 9 out of 9 (Fig. 1). Plasma viraemia is usually detectable after 14 days, occasionally 21 days (Supplementary Information), with seroconversion after 14–35 days. Inhibitor-treated animals remaining plasma RNA-negative and antibody-negative for 70 days after challenge are considered protected^{14,15}.

We had previously tested CMPD167 as a 1 mM (0.6 mg ml⁻¹) solution in 2.5% hydroxymethylcellulose (HMC) gel for protection against the related SHIV-162P4 virus⁸. Only 2 out of 11 macaques remained uninfected, but reduced viraemia in infected animals suggested partial protection⁸. We adjusted the gel pH to 6.0, which allowed us to now dissolve CMPD167 at 5 mM. At pH 6.0, HMC was neither inhibitory nor inflammatory. It was used in all the following studies.

Eight out of ten macaques receiving 5 mM CMPD167 remained uninfected after SHIV-162P3 challenge (Fig. 1). Two more monkeys (of two tested) given 5 mM CMPD167 in combination with soluble mannan (50 mg ml⁻¹) were also uninfected. They had probably been protected by CMPD167, because 2 out of 2 animals given mannan alone became infected (Fig. 1). In addition, 2 out of 3 monkeys that received low-dose CMPD167 (1 mM) were protected against SHIV-162P3 (Fig. 1). Overall, 12 out of 15 macaques receiving CMPD167 (1–5 mM) with or without mannan were protected against SHIV-162P3 infection.

Our studies with CMPD167, and other reports, indicate that protection *in vivo* can only be achieved using inhibitor concentrations orders of magnitude greater than are active *in vitro* (that is, in the millimolar range as compared with the nanomolar range)^{14–16}. We therefore tested BMS-378806 at the highest concentration soluble in HMC, estimated as 5.5 mM (2.3 mg ml⁻¹). Six out of eight animals remained uninfected after SHIV-162P3 challenge (Fig. 1).

We tested ever-increasing concentrations of C52L, initially using SHIV-162P4 before later changing to SHIV-162P3. Three out of five animals given 1.5 mM C52L remained uninfected after SHIV-162P3 challenge (Fig. 1). However, taking the SHIV-162P3 and SHIV-162P4 challenges together, C52L provided ~50% protection over a wide concentration range (0.05–1.5 mM, ~0.3–9 mg ml⁻¹), with no clear dose–response relationship (Fig. 1; see also Supplementary Fig. 1). Either C52L alone cannot be fully protective, or we could not achieve an effective concentration much beyond ~0.05 mM simply by adding more peptide to the gel.

CMPD167 (5 mM), BMS-378806 (5.5 mM) and C52L (1.5 mM) were each significantly protective compared to controls

¹Tulane National Primate Research Center, Covington, Louisiana 70433, USA. ²Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York 10021, USA. ³St George's, University of London, London SW17 0RE, UK. ⁴Department of Biochemistry, Weill Medical College of Cornell University, New York, New York 10021, USA. ⁵Bristol-Myers Squibb Pharmaceutical Institute, Wallingford, Connecticut 06492, USA. ⁶Merck Research Laboratories, Rahway, New Jersey 07065, USA.

Table 1 | Inhibition of virus infection by single inhibitors *in vitro*

Virus	Cells	50% inhibitory concentration (IC ₅₀ ; nM)*		
		CMPD167	BMS-378806	C52L
HIV-1 _{BaL}	Cervical explants (n = 2)	5.4 ± 0.48	31 ± 5.2	16 ± 6.5
HIV-1 _{BaL}	Macrophages (n = 3)	0.69 ± 0.16	6.2 ± 3.3	2.8 ± 0.96
SHIV-162P3	Tzm-bl (n = 5–6)	0.22 ± 0.11	5.5 ± 0.98	8.1 ± 5.2
SHIV-162P3	Human PBMCs† (n = 1)	0.10	11	8.6
SHIV-162P3	Macaque PBMCs (n = 1)	0.14	6.9	15

*Values are means from the number of experiments indicated for each cell type (±s.d.).

†Pooled PBMCs from four human donors were used.

($P = 0.00060$, 0.0023 and 0.027 , respectively). In total, 21 out of 28 macaques receiving one inhibitor 30 min before challenge (in two cases also with mannan) were protected ($P = 0.000092$). Viral loads in the infected test animals were also analysed (Supplementary Information).

The rationale for inhibitor combinations is as strong for protection as for treatment^{17,18}. As a first step towards a practical microbicide formulation, we tested inhibitor pairs. We used CMPD167 and BMS-378806 at reduced concentrations of 1 mM and 2 mM (0.8 mg ml^{-1}), respectively, to see whether dose reductions might be possible with two inhibitors. C52L was used at 1.5 mM, because its protective effect did not seem dose dependent (Supplementary Fig. 1). The 33% infection rate observed in the three monkeys receiving only 1 mM CMPD167 before SHIV-162P3 challenge falls between that seen when using either compound alone at the higher concentration (20–25%) and the 50% infection rate observed when CMPD167 (1 mM) was combined with BMS-378806 (2 mM) (Fig. 1). These differences were not statistically significant. We have not tested BMS-378806 by itself at 2 mM, but, by inference, we would not expect it to be highly protective (that is, not >50%) at this concentration. A practical combination of CMPD167 with BMS-378806 may require high concentrations of both compounds, consistent with their additive, and not synergistic, inhibitory effects *in vitro* (Supplementary Information).

Substantial protection was observed when C52L (1.5 mM) was combined with either CMPD167 (1 mM) or BMS-378806 (2 mM), with 1 out of 6 and 0 out of 6 macaques becoming infected, respectively (Fig. 1). The infection rate for the combination of

BMS-378806 and C52L (0 out of 6) was lower than with C52L alone (2 out of 5, $P = 0.18$); for the CMPD167 plus C52L combination (1 out of 6) it was marginally lower than for either compound alone at the same concentration (1 out of 3, $P = 0.58$ and 2 out of 5, $P = 0.42$). These differences were not statistically significant (Fig. 1). In addition, 2 out of 2 macaques receiving high-dose CMPD167 (5 mM) plus C52L (1.5 mM) were protected (Fig. 1). Thus, there are indications that C52L increases the protective capacities of CMPD167 and BMS-378806, consistent with the synergy observed with these combinations *in vitro* (Supplementary Information).

Different inhibitor pairs were protective compared to control animals (9 out of 9 infected). For CMPD167 (1 mM) plus C52L (1.5 mM), $P = 0.0020$; for CMPD167 (1 mM) plus BMS-378806 (2 mM), $P = 0.044$; for BMS-378806 (2 mM) plus C52L (1.5 mM), $P = 0.00020$. Overall, 16 out of 20 macaques given two inhibitors were protected ($P = 0.000071$). In addition, we tested the triple combination of CMPD167 (5 mM), BMS-378806 (5.5 mM) and C52L (1.5 mM) in three animals, all of which were protected ($P = 0.0045$).

It would be desirable if a microbicide could protect women for several hours after application². Although the local concentrations of topically applied compounds will decrease over time, small-molecule CCR5 inhibitors remain receptor-bound for prolonged periods *in vitro*^{19,20}, and there are indications of sustained inhibitory effects *in vivo*^{8,20}. We therefore delayed SHIV-162P3 challenge for 2 h, 6 h or 12 h after adding CMPD167 (5 mM), the macaques being conscious and mobile during the intervening period. The infection rates at each time delay were 1 out of 3, 3 out of 6 and 2 out of 3, respectively. When BMS-378806 (5.5 mM) was added 2 h or 6 h before challenge, 1 out of 2 and 1 out of 3 animals became infected (Fig. 1). We have not yet tested C52L or inhibitor combinations against delayed challenge. Overall, we conclude that sustained protection may be possible.

The above studies show that we can consistently protect macaques against a vaginally inoculated R5 SHIV, using three different, specific entry inhibitors delivered vaginally. In total, 40 out of 51 animals (78%) given one, two or three inhibitors 0.5 h before challenge were protected. Compared to controls (0 out of 9 uninfected), the difference is highly significant ($P = 0.000011$, Fig. 1). In our high-dose model, a single virus challenge is given after the vaginal epithelium has been thinned by progesterone treatment^{8,12,14,15,21}. The challenge dose contained 5×10^7 RNA copies, 300 TCID₅₀ (50% tissue culture infectious dose) and >2 animal infectious units. Some comparisons are useful: In the 'low-dose' model, the challenge dose is 2 or 10 TCID₅₀, given on multiple occasions²². The highest recorded amount of HIV-1 RNA in human semen is $\sim 2 \times 10^6$ RNA copies²³, and the highest estimate of the risk of a woman acquiring infection during a single coital act is 1 out of 10 (ref. 24). We estimate that translates to an exposure to ~ 0.1 human infectious units. Hence, the high-dose challenge model provides a stringent test of vaginally applied compounds. Robust protection is encouraging for development of a microbicide for women.

The concentrations of all three compounds required for consistent protection (1–10 mM) are much higher than are inhibitory *in vitro* (1–10 nM). Similar differentials were seen in studies involving

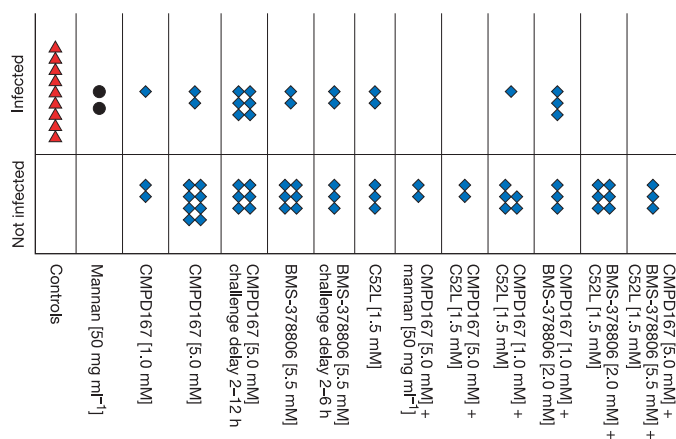


Figure 1 | Vaginally delivered inhibitors can protect macaques against vaginal challenge with SHIV-162P3. Each symbol represents a single macaque, grouped by infection status after challenge. Control animals are shown by red triangles and received HMC gel, pH 6.0; 9 out of 9 became infected. Two monkeys received soluble mannan (50 mg ml^{-1}) to try to block attachment to receptors such as DC-SIGN¹⁷ (black circles). The 68 animals given CMPD167, BMS-378806 or C52L, alone or in combination, are depicted by blue diamonds. The virus challenge was 30 min after inhibitor delivery, except for animals denoted as 'CMPD167 [5.0 mM] challenge delay 2–12 h' and 'BMS-378806 [5.5 mM] challenge delay 2–6 h'.

SHIV-162P or SHIV-89.6P and b12, PSC-RANTES or cyanovirin-N, compounds that interfere with the fusion process in different ways^{14–16}. The general requirement for high inhibitor concentrations *in vivo* may reflect the need to shield adequately all exposed vaginal surfaces from a high titre inoculum for a significant period of time. It is possible that lower concentrations might be sufficient to protect women, but we believe it would be an imprudent assumption. Men undergoing primary infection can be highly infectious, and are considered to be the driving force behind the rapid local spread of HIV-1 in a population^{23,24}. Countering transmission during primary infection may be both critically important and difficult²⁴. We therefore favour using multiple inhibitors at the highest practical concentrations that can be suitably formulated.

Attacking multiple targets could help combat HIV-1 sequence diversity and minimize the transmission of variants resistant to any single inhibitor. It does not matter whether a topical inhibitor of viral entry targets the virus (for example, BMS-378806 or C52L) or the cell (for example, CMPD167); either method can work. BMS-378806 and CMPD167 provided additive inhibition *in vitro*, and C52L acts synergistically with both compounds, probably via established molecular mechanisms^{18,25}. Its presence, even at relatively low concentrations, might be advantageous. A practical microbicide must be cheap to manufacture². The cost of making conventional small molecules and a bacterially expressed peptide might not be unreasonable. In contrast, proteins, monoclonal antibodies and synthetic peptides may be prohibitively expensive in the millimolar amounts likely to be needed per application.

It should be possible to identify better analogues of the particular compounds we have studied here. As with most drug candidates, CMPD167 and BMS-378806 represent structural compromises between potency *in vitro* and biological availability *in vivo* after oral administration. For vaginal delivery, potency and effective residence times are paramount, whereas oral biological availability is irrelevant. We have already identified structural variants of CMPD167 and BMS-378806 that are up to 50-fold more active against SHIV-162P3 *in vitro* (our own unpublished results). We cannot address issues relating to breadth of protection using a single challenge virus, and synergy studies *in vivo* are impractical, so we will need to rely on susceptibility patterns in cell-based assays. Although CMPD167 and C52L inhibit HIV-1 isolates from multiple subtypes, BMS-378806 activity is greatest against subtype B strains^{3,4}. However, newer analogues are more broadly active^{3,4}. BMS-378806 and C52L can inhibit X4 viruses, but CMPD167, of course, cannot^{3,4,6,9}. Transmission of X4 viruses is rare, but can occur^{2,13}. Thus, we will soon test whether a small-molecule CXCR4 inhibitor can protect against an X4 SHIV.

Safety in women, particularly during prolonged use, can only be determined in future clinical trials. However, by choosing compounds similar to ones with acceptable safety profiles in humans, we hope to avoid major concerns. We observed no vaginal irritation or inflammation in short-term macaque studies. Appropriate formulations designed to optimize inhibitor solubility, absorption and distribution must now be sought, to maximize both efficacy and acceptability. It is also possible that microbicide use by unknowingly infected women might lower the risk of transmission to male partners.

The different inhibitors probably confer protection because they prevent the incoming virus from fusing with CD4- and CCR5-expressing cells within or near the vaginal epithelium that are in physical and temporal range of topically applied compounds, either by binding to the virus itself or to the CCR5 co-receptor. Such cells are infected rapidly after virus inoculation into the vagina²⁶. It has been suggested that virions can be transported within migratory dendritic cells to draining lymphoid tissues, days later and centimetres away^{2,27}. However, this pathway may be dependent upon localized infection of dendritic cells, and therefore also sensitive to topically applied entry inhibitors^{27,28}.

An effective HIV-1 prevention strategy may require a combination of approaches. Inhibitors could be given both orally and vaginally, if costs allow; orally delivered antiviral drugs could supplement the protective capacity of partially effective vaccines; and vaccine antigens could be delivered mucosally, formulated in a microbicide.

METHODS

Inhibitors. BMS-378806 was synthesized at Bristol-Myers Squibb Inc²⁹, and CMPD167 at Merck Research Labs⁸. The C52L peptide NHTTWMEWDREIN-NYTSLIHSLIEESQNLQEKNEQELLELDKWASLWNWFNIK was expressed in *Escherichia coli* and purified by reverse-phase high-performance liquid chromatography as described elsewhere³⁰.

Inhibition of virus replication *in vitro*. Inhibition assays using PBMCs from one macaque and pooled PBMCs from four human donors, or using human MDMs and human cervical tissue explants, have all been described previously^{8,14,17}. Cervical tissues were obtained from pre-menopausal women undergoing planned therapeutic hysterectomy in the absence of any cervical pathology at St George's, University of London, UK (written consent was obtained from all tissue donors, supervised by the local Research Ethics Committee). Tissues were dissected into 3 × 3 mm explants and cultured in 200 µl of supplemented RPMI 1640 in a 96-well plate¹⁷. Human mononuclear cells were isolated from buffy coat preparations (National Blood Transfusion Service, UK) by density gradient separation (Histopaque-1077, Sigma). After adhesion onto plastic for 2 h at 37 °C, PBMCs were further depleted of lymphocytes by extensive washing in culture medium. Adherent monocytes were cultured in 200 µl of RPMI 1640 containing 10% human serum, 5% FCS in a 96-well plate for 7 days.

Explants or MDMs were incubated with serially diluted inhibitors, alone or in combination, for 1 h at 37 °C. HIV-1_{BAL} (2,000 TCID₅₀ for explants, 200 for MDMs) was then added for 3 h before unbound virus was removed by washing. Serially diluted inhibitors were then added back to the cultures. Supernatants were harvested every 3–4 days and stored at –80 °C. Viral replication was determined at day 14 by p24 enzyme-linked immunosorbent assay (ELISA) for explants or by reverse transcription assay for MDMs. The p24 antigen ELISA was performed according to the manufacturer's protocol (Beckman Coulter). Reverse transcription activity was determined by incorporation of ³²P-dTTP. Fifty per cent inhibitory concentration (IC₅₀) values were calculated by sigmoid-curve fitting in Prism 4 (Graphpad). To detect synergy, additivity or antagonism, median-effect-dose values and combination indices (CI) were calculated by use of the Chou-Talalay equations, as included in Calcsyn (Biosoft). All data points are derived from triplicate cultures.

Macaque studies. Normal cycling, adult (4–12 yr old) female rhesus macaques (*Macaca mulatta*) were used. All studies adhered to the Guide for the Care and Use of Laboratory Animals, prepared by the National Research Council, NIH, and with the Guidelines of the Tulane University Institutional Animal Care and Use Committee (IACUC). The animals were treated with a single 30-mg intramuscular injection of depo-medroxyprogesterone acetate (Depo-Provera) 28–34 days before experiments, to synchronize their menstrual cycles and thin the vaginal mucosa, which increases their susceptibility to vaginal transmission²¹. They were then sedated with Telazol and placed in ventral recumbency with their hips elevated. Inhibitors were delivered atraumatically into the vaginal vault using a pliable French catheter 30 min before virus challenge. Usually, the inhibitors were formulated in 5 ml of 2.5% HMC gel, but phosphate-buffered saline (PBS) was used in one control monkey. The challenge virus was 300 TCID₅₀ of SHIV-162P3 (obtained from the NIH AIDS Research and Reference Reagent Program) in 1 ml of RPMI 1640 medium^{12,15}. In the delayed challenge experiments, the macaques were maintained in ventral recumbency for 30 min, returned to their cages for 2–12 h, then re-anaesthetized and vaginally challenged as described above.

Blood was collected for analysis in EDTA tubes every week after challenge for 28 days and every other week thereafter through to 70 days after infection. In all experiments, protection is defined as the complete absence of plasma viraemia (<125 RNA copies per ml) at any weekly time point from 7 to 70 days after challenge, and the lack of seroconversion to SIV/SHIV antibodies throughout this period, as previously described^{14,15}. The proportions of infected and uninfected animals were analysed using Fisher's exact test as included in the Prism program (Graphpad)⁸.

Received 26 May; accepted 21 July 2005.

Published online 30 October 2005.

1. Klausner, R. D. *et al.* Medicine. The need for a global HIV vaccine enterprise. *Science* 300, 2036–2069 (2003).

2. Shattock, R. A. & Moore, J. P. Inhibiting HIV-1 sexual transmission. *Nature Rev. Microbiol.* **1**, 25–34 (2003).
3. Lin, P. F. *et al.* A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding. *Proc. Natl Acad. Sci. USA* **100**, 11013–11018 (2003).
4. Guo, Q. *et al.* Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120–CD4 interactions. *J. Virol.* **77**, 10528–10536 (2003).
5. Shen, D. M. *et al.* Antagonists of human CCR5 receptor containing 4-(pyrazolyl)piperidine side chains. Part 2: Discovery of potent, selective, and orally bioavailable compounds. *Bioorg. Med. Chem. Lett.* **14**, 941–945 (2004).
6. Lu, M. Stabilizing peptides and their use in the preparation of stabilized HIV inhibitors. World Intellectual Property Organization Patent WO-04/106364A1 (2004).
7. Seibert, C. & Sakmar, T. P. Small-molecule antagonists of CCR5 and CXCR4: A promising new class of anti-HIV-1 drugs. *Curr. Pharmacol. Des.* **10**, 2041–2062 (2004).
8. Veazey, R. S. *et al.* Use of a small-molecule CCR5 inhibitor in macaques to treat simian immunodeficiency virus infection and prevent simian-human immunodeficiency virus infection. *J. Exp. Med.* **198**, 1551–1562 (2003).
9. Wolinsky, S. M. *et al.* Effect of a CCR5 inhibitor on viral loads in macaques dual-infected with R5 and X4 primate immunodeficiency viruses. *Virology* **328**, 19–29 (2004).
10. Hanna, G. *et al.* Antiviral activity, safety, and tolerability of a novel, oral small-molecule HIV-1 attachment inhibitor, BMS-488043, in HIV-1-infected subjects. Abstr. 141, *11th Conf. Retroviruses Opportunistic Infect.* (http://www.retroconference.org/Search_Abstract_2004/AbstractSearch.aspx) (2004).
11. Matthews, T. *et al.* Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nature Rev. Drug Disc.* **3**, 215–221 (2004).
12. Harouse, J. M. *et al.* Mucosal transmission and induction of simian AIDS by CCR5-specific simian/human immunodeficiency virus SHIV_{SF162P3}. *J. Virol.* **75**, 1990–1995 (2001).
13. Moore, J. P., Kitchen, S. G., Pugach, P. & Zack, J. A. The CCR5 and CXCR4 coreceptors—central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS Res. Hum. Retroviruses* **20**, 111–126 (2004).
14. Veazey, R. S. *et al.* Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120. *Nature Med.* **9**, 343–346 (2003).
15. Lederman, M. M. *et al.* Prevention of vaginal SHIV transmission in rhesus macaques through inhibition of CCR5. *Science* **306**, 485–487 (2004).
16. Tsai, C. C. *et al.* Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res. Hum. Retroviruses* **20**, 11–18 (2004).
17. Hu, Q. *et al.* Blockade of attachment and fusion receptors inhibits HIV-1 infection of human cervical tissue. *J. Exp. Med.* **199**, 1065–1075 (2004).
18. Tremblay, C. Effects of HIV-1 entry inhibitors in combination. *Curr. Pharm. Des.* **10**, 1861–1865 (2004).
19. Watson, C., Jenkinson, S., Kazmierski, W. & Kenakin, T. P. The CCR5 receptor-based mechanism of action of 873140, a potent allosteric non-competitive HIV entry-inhibitor. *Mol. Pharmacol.* **67**, 1268–1282 (2005).
20. Sparks, S. *et al.* Prolonged duration of CCR5 occupancy by 873140 in HIV-negative and HIV-positive subjects. Abstr. 77, *12th Conf. Retroviruses Opportunistic Infect.* (http://www.retroconference.org/Search_Abstract_2005/AbstractSearch.aspx) (2005).
21. Marx, P. A. *et al.* Progesterone implants enhance SIV vaginal transmission and early virus load. *Nature Med.* **2**, 1084–1089 (1996).
22. Otten, R. A. *et al.* Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *J. Infect. Dis.* **191**, 164–173 (2005).
23. Dyer, J. R. *et al.* High levels of human immunodeficiency virus type 1 in blood and semen of seropositive men in sub-Saharan Africa. *J. Infect. Dis.* **177**, 1742–1746 (1998).
24. Wawer, M. J. *et al.* Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J. Infect. Dis.* **191**, 1403–1409 (2005).
25. Reeves, J. D. *et al.* Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proc. Natl Acad. Sci. USA* **99**, 16249–16254 (2002).
26. Miller, C. J. *et al.* Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J. Virol.* **79**, 9217–9227 (2005).
27. Pope, M. & Haase, A. T. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nature Med.* **9**, 847–852 (2003).
28. Ketas, T. J. *et al.* Entry inhibitors SCH-C, RANTES, and T-20 block HIV type 1 replication in multiple cell types. *AIDS Res. Hum. Retroviruses* **19**, 177–186 (2003).
29. Wang, T. *et al.* Discovery of 4-Benzoyl-1-[(4-methoxy-1H-pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2-(R)-methylpiperazine (BMS-378806): A novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. *J. Med. Chem.* **46**, 4236–4239 (2003).
30. Ji, H., Bracken, C. & Lu, M. Buried polar interactions and conformational stability in the simian immunodeficiency virus (SIV) gp41 core. *Biochemistry* **39**, 676–685 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. LeBlanc, M. Dodd and T. Williams for technical support, and we appreciate receiving advice from M. Pope. This work was supported by an NIH grant, and by an Unrestricted Infectious Diseases Award from the Bristol-Myers Squibb Foundation to J.P.M.

Author Contributions R.S.V. was responsible for designing, organizing and executing the macaque studies, which were carried out by J.D. P.A.M. provided advice on the macaque model and supervised the performance of serology studies on samples from the test monkeys. P.J.K. contributed to the design of the monkey and cell culture experiments and performed all the mathematical and statistical analyses. S.M.S. was responsible for *in vitro* studies of inhibitor combinations, and T.J.K. for infection-inhibition experiments with human and macaque PBMCs. Q.H. carried out the experiments involving cervical tissue explants and macrophages (including combination studies), under the supervision of R.J.S., who was also involved in the design and interpretation of the macaque studies. M.L. synthesized the C52L peptide and advised on its use. R.J.C. provided BMS-378806 and advised on its use. M.S.S. provided CMPD167 and advised on its use. J.P.M. made a lot of phone calls and sent out many emails.

Author Information The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence should be addressed to J.P.M. (jpm2003@med.cornell.edu); see Supplementary Information for details about requests for materials.

A protein interaction network of the malaria parasite *Plasmodium falciparum*

Douglas J. LaCount^{1,2*}, Marissa Vignali^{2*}, Rakesh Chettier³, Amit Phansalkar³, Russell Bell³, Jay R. Hesselberth², Lori W. Schoenfeld^{1,2}, Irene Ota³, Sudhir Sahasrabudhe³, Cornelia Kurschner³, Stanley Fields^{1,2} & Robert E. Hughes^{3†}

Plasmodium falciparum causes the most severe form of malaria and kills up to 2.7 million people annually¹. Despite the global importance of *P. falciparum*, the vast majority of its proteins have not been characterized experimentally. Here we identify *P. falciparum* protein–protein interactions using a high-throughput version of the yeast two-hybrid assay that circumvents the difficulties in expressing *P. falciparum* proteins in *Saccharomyces cerevisiae*. From more than 32,000 yeast two-hybrid screens with *P. falciparum* protein fragments, we identified 2,846 unique interactions, most of which include at least one previously uncharacterized protein. Informatic analyses of network connectivity, coexpression of the genes encoding interacting fragments, and enrichment of specific protein domains or Gene Ontology annotations² were used to identify groups of interacting proteins, including one implicated in chromatin modification, transcription, messenger RNA stability and ubiquitination, and another implicated in the invasion of host cells. These data constitute the first extensive description of the protein interaction network for this important human pathogen.

The 80% AT content of the *P. falciparum* genome³ hinders protein expression in heterologous systems⁴ and limits both conventional biochemical approaches and comprehensive analyses of this organism's proteins. We overcame this problem by applying a yeast two-hybrid approach that makes use of protein fusions carrying three elements: the Gal4 DNA-binding or activation domain, a random fragment of a *P. falciparum* protein, and an enzyme that allows the growth of auxotrophic yeast deleted for the cognate gene (Supplementary Fig. 1). This procedure was designed to select only those yeast transformants that contained plasmids encoding in-frame and expressed fragments of *P. falciparum* proteins. In addition, this procedure enabled us to generate libraries of DNA-binding domain fusions ('baits') from which randomly chosen transformants could be individually screened in parallel against a library of activation domain fusions ('preys'). Our libraries were derived from RNA isolated from mixed intra-erythrocytic-stage parasites—the stage responsible for pathogenesis in humans—and thus lack genes expressed exclusively in the liver, gametocyte or mosquito stages. Sequence analysis of the inserts from 4,456 DNA-binding domain fusions identified more than 2,000 non-overlapping gene fragments representing 1,295 different *P. falciparum* genes expressed throughout the intraerythrocytic cycle (Supplementary Fig. 2a), indicating that our libraries are complex. Because relatively small fragments of *P. falciparum* genes (about 450 base pairs on average) were cloned, narrowly defined protein–protein interaction domains could be identified.

Coverage of the proteome was obtained by performing more than

32,000 yeast two-hybrid screens, of which 11% yielded positives in which the identities of both interacting protein fragments were determined (Supplementary Table 1). Because the complete set of putative interactions contains both true and false positives, we first sought to eliminate the most obvious class of false positives, namely those protein fragments with many partners, which seem to be 'promiscuous' in the two-hybrid assay. Indeed, although the vast majority of fragments identified relatively few partners, some interacted with many partners (up to a maximum of 207). To identify promiscuous protein fragments, we sequentially applied *k*-means clustering analysis (with *k* = 2) to prey and bait fragments to define two populations based on the number of interacting partners. This approach identified 13 promiscuous prey fragments with more than 31 partners, and 28 promiscuous bait fragments with more than 25 partners (Supplementary Table 2), resulting in the removal of 2,155 interactions involving these fragments from the data set. Because the remaining interactions are listed as pairs of interacting proteins rather than fragments, the total number of partners for a given protein can exceed the thresholds used to remove promiscuous fragments if the protein contains multiple non-promiscuous fragments. Although this analysis removes a significant number of non-specific interactions, other classes of false positive, including promiscuous fragments that resulted in fewer interactions than our threshold values and two-hybrid pairs resulting from mutations in the plasmids or reporter strain, remain in the data set. False positives have been noted in other high-throughput two-hybrid data sets (for example those for proteins of *Drosophila*^{5,6}), and can make up a substantial portion of the reported interactions.

The 2,846 unique pairwise interactions that remain constitute a core data set (Supplementary Table 3, also available from the PlasmoDB (<http://www.plasmodb.org/>) and BIND (<http://bind.ca/>) databases) that includes 25% of the predicted *P. falciparum* proteins and forms a highly interconnected, scale-free network⁷ containing 1,267 proteins linked by 2,823 interactions. An additional 41 proteins are present in small groups of one or two interactions. All categories of proteins seem to have been sampled approximately in proportion to their representation in the genome (Supplementary Fig. 2b). The core data set includes 23 interactions that were previously observed either in *Plasmodium* or between orthologous proteins (Supplementary Table 4). In all, 82% of the interactions include at least one protein annotated as 'hypothetical', and 33% of the interactions include two hypothetical proteins. The difficulties in expressing *P. falciparum* proteins in heterologous systems precluded experimental confirmation of the interactions by another method. To circumvent this problem, we used several independent bioinformatic analyses to uncover biologically interesting regions of the network.

¹Howard Hughes Medical Institute, ²Departments of Genome Sciences and Medicine, University of Washington, Box 357730, Seattle, Washington 98195, USA. ³Prolexis Pharmaceuticals, Inc., 2150 West Dauntless Avenue, Salt Lake City, Utah 84111, USA. †Present address: Buck Institute, Novato, California 94945, USA.

*These authors contributed equally to this work.

Because high degrees of local network interconnectivity can identify sets of functionally related proteins⁸, we surveyed the network for groups of proteins with a greater number of connections than would be expected by chance. We parsed the network into 1,308 primary subnetworks containing a protein, its direct binding partners and all interactions between them, and calculated a connectivity coefficient (defined as the number of interactions divided by the number of proteins). A comparison of the distribution of connectivity coefficients present among the experimentally observed subnetworks with those derived from randomized subnetworks of equal size showed an enrichment for highly connected subnetworks in the real data (Fig. 1a); 96 subnetworks showed a higher degree of interconnectivity than expected by chance ($P \leq 0.05$) when compared with the mean of the connectivity coefficients from 100 randomized subnetworks of the same size (Supplementary Table 5). Several of these 96 subnetworks shared a common set of interactions that revealed a region of the interaction network with a high degree of

interconnectivity. This region overlaps the complex with the highest connectivity in the data set as defined by MCODE⁹, an algorithm that identifies densely connected areas of protein interaction networks (Supplementary Fig. 3, complex 1). On the basis of these analyses, we identified a group of interacting proteins likely to integrate chromatin modification, transcriptional regulation, mRNA stability and ubiquitination (Fig. 1b). Whereas 11% of the interactions in the whole data set were observed in two or more independent experiments, more than 40% of the interactions in this group were observed in multiple independent experiments.

This set of interacting proteins is centred on PF08_0034, the *P. falciparum* orthologue of the yeast histone acetyltransferase Gcn5. The interactions established by PfGcn5 are mediated by an amino-terminal extension that is absent from the yeast Gcn5 (ref. 10), indicating that this group might represent a *Plasmodium*-specific pathway that regulates gene expression. Other potential chromatin-modifying proteins in this group include PFF1440w, a protein containing a PHD domain, bromodomain and SET domain that potentially recognizes acetylated nucleosomes and acts as a histone methyltransferase^{11,12}, and PFF1470c, an orthologue¹³ of yeast DNA Pol2 (Pole), which is involved in DNA replication and chromatin silencing at telomeres¹⁴. Two putative transcription factors are also present in this group: PF11_0241, which contains a Myb DNA-binding domain (and directly interacts with both PF08_0034 and PFF1440w), and PF10_0075, which has a DNA-binding AT-hook and a BTB/POZ domain, which is found in some transcription factors involved in the recruitment of histone deacetylase complexes¹⁵. These interactions indicate that chromatin-modifying complexes might be targeted to specific regions in the genome to regulate transcription and are of particular significance given the apparently unique features of gene expression of the parasite^{16,17} and the dearth of recognizable transcription factors encoded by its genome^{18,19}. The presence of three ubiquitin metabolism proteins (the RING finger and forkhead-associated (FHA)-domain protein PFL0275w, the HECT-domain protein PFF1365c and the UCH-domain protein PFI0225w) indicates that ubiquitination might be involved in regulating the stability or activity of proteins in this group; Gcn5-containing HAT complexes can deubiquitinate histones²⁰. This group also contains MAL8P1.104, the *P. falciparum* orthologue of Caf1, the major mRNA deadenylase in yeast and a member of the Ccr4–Not complex. Indeed, this group seems analogous to Ccr4–Not, which also integrates transcription regulation, chromatin modification, ubiquitination and RNA stability²¹. Other interactions implicated in *Plasmodium* nucleic acid metabolism^{18,19} are shown in Supplementary Fig. 4.

Interacting proteins from *S. cerevisiae* tend to have similar mRNA abundance profiles²²; positively correlated mRNA expression is therefore generally taken as an indication that a putative interaction is more likely to be real. However, this correlation is not as evident in the large-scale *C. elegans* protein interaction data²³, and two genes need not share similar expression profiles for their proteins to be present in the cell at the same time. For example, one gene may be expressed constitutively whereas the other is induced under certain conditions, or their proteins may have different half-lives. Indeed, the time of maximal accumulation for a substantial portion of *P. falciparum* proteins is shifted relative to the time of maximum mRNA abundance²⁴. To examine the relationship between *P. falciparum* protein interaction data and mRNA abundance, we compared our core data set to data from two genome-scale gene expression studies that addressed the timing of mRNA accumulation during the *P. falciparum* life cycle^{16,17}. We calculated Pearson correlation coefficients (PCCs) for each protein pair and averaged these values for proteins with more than one partner. We identified 82 proteins with average PCCs significantly higher than expected compared with mean PCCs from 100 randomizations in which the identity of the partners was changed but the total number of partners remained constant ($P \leq 0.05$; Supplementary Table 6). Several of

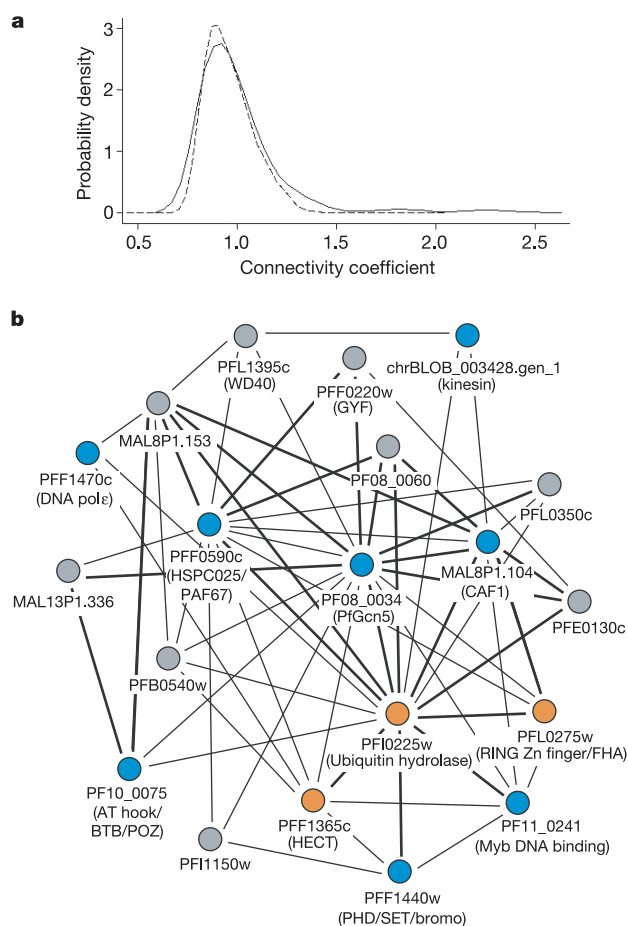


Figure 1 | Connectivity analysis. **a**, Distribution of connectivity coefficients from experimental (solid lines) and randomized (dashed line) subnetworks containing five or more proteins (a central node plus four or more partners). **b**, The most highly connected region of the *P. falciparum* protein interaction network. The graph shows interactions (lines) between proteins (circles) involved in transcription or chromatin metabolism (blue) and ubiquitin metabolism (orange); proteins with no additional supporting evidence linking them to these processes are shown in grey. Thin lines indicate interactions observed in a single yeast two-hybrid experiment; thick lines show interactions found in reciprocal orientation or in two or more independent yeast two-hybrid screens. Systematic gene names are as given in PlasmoDB; when available, common names, putative functions and domains are shown in parentheses. Only proteins with two or more interactions with other group members are shown; additional partners for these proteins are listed in Supplementary Table 3.

associated with protein domains or GO annotations (Supplementary Tables 8 and 9). We identified several subnetworks with an enrichment of RNA recognition motifs (RRM; Fig. 3a), indicating that the proteins in these subnetworks might be involved in RNA processing or splicing. For example, PF07_0082, a hypothetical protein, interacts with three proteins that have RRM domains and a fourth that might be an orthologue of the *Drosophila* splicing factor suppressor of white apricot. PFE0750c, a potential orthologue of the *S. cerevisiae* splicing factor Cwc2, interacts with four putative RNA-binding proteins and an orthologue of the yeast splicing protein Prp40. PFI1715w binds to four putative RNA-processing proteins, including orthologues of yeast splicing factors Prp4 and Prp18, and human alternative splicing factor 2; the putative *P. falciparum* Prp4 protein in turn binds to the *P. falciparum* orthologue of yeast Prp9 (PFI1215w, not shown). PFI1715w also interacts with three proteins bearing SNF2 helicase domains (Fig. 3b). Given that proteins containing SNF2 domains are involved in chromatin remodelling, PFI1715w might provide a link between gene expression and splicing in *P. falciparum*. In some cases, domain and GO annotation enrichment indicate unexpected alternative functions for *P. falciparum* proteins. We found an orthologue of the yeast nucleosome assembly protein Nap1 (PFL0185c) in interactions with nine ribosomal proteins (Fig. 3c) and a protein involved in ribosome biogenesis (PFI1_0259, Rrs1; not shown), possibly indicating a role for PFL0185c in ribosome assembly or translation. Similarly, a putative subunit of the N-terminal acetyltransferase complex (PFL2120w) interacts with the cytoplasmic tails of four PfEMP-1 (*P. falciparum* erythrocyte membrane protein-1) proteins (Fig. 3d) and two other proteins that are known or predicted to be exported to the host cell cytoplasm (MAL7P1.170 and PFA0110w, not shown), indicating that acetylation (or binding to PFL2120w) might be involved in protein trafficking to this compartment. Other examples of domain enrichment implicate uncharacterized *P. falciparum* proteins in protein folding (Fig. 3e) and a possible protein kinase signalling cascade (Fig. 3f).

Last, we compared our set of interactions with the more than 150 parasite-derived proteins known or predicted to be exported into the host cell cytoplasm^{28–30}. These proteins extensively modify infected red blood cells, establishing a new secretory system in the cytoplasm and generating knob-like structures on the cell surface. We identified 15 interactions among 19 exported proteins, which might provide insight into the structure of parasite-mediated modifications of the host cell (Supplementary Fig. 7).

The data set of putative protein interactions described here greatly exceeds the number of previously known interactions for *P. falciparum* and provides the basis for focused experiments on a variety of biological processes. As this network reflects pathways and processes of the parasite, this information should be relevant both to our understanding of the basic biology of the organism and to the discovery of new drug and vaccine targets. Although the difficulties in working with this organism currently preclude the types of experimental validation that are available in model organisms, we expect that the accumulation and integration of large-scale gene and protein expression studies and protein interaction data sets will continue to provide informatics-based approaches towards understanding this human parasite.

METHODS

Bait and prey construction. Complementary DNA was generated from poly(A)⁺ RNA isolated from mixed-staged (strain 3D7)-infected erythrocytes (a gift from K. Ganesan and P. Rathod) and inserted between the Gal4 transcriptional activation domain and the *Schizosaccharomyces pombe* URA4 coding region of pOAD.102 (prey plasmid) or the Gal4 DNA-binding domain and the *S. cerevisiae* MET2 coding region of pOBD.111 (bait plasmid). Yeast transformed with bait or prey plasmids were plated on medium lacking uracil (prey) or methionine (bait) to select for transformants expressing the markers fused to the cDNA inserts. Additional information about the plasmids, yeast strains and library construction can be found in Supplementary Methods.

Yeast two-hybrid process. Individual bait colonies were picked at random and clonally expanded in liquid medium in 96-well plates. Aliquots from the prey libraries were added to each well; mating occurred overnight. Matings were plated on medium that selected for the mating event, the expression of the auxotrophic markers fused to the cDNA inserts, and the activity of the metabolic reporter genes *ADE2* and *HIS3*. The cDNA inserts from yeast that grew on this selection medium were amplified by polymerase chain reaction and then sequenced. The identities of inserts were determined by querying the sequences against the annotated *P. falciparum* genes in PlasmoDB version 4.0 and the genome sequences from PlasmoDB version 3.3 that were excluded from version 4.0. Additional details are provided in Supplementary Methods.

Removal of false-positive bait and prey fragments. Activation and DNA-binding domain inserts were treated as protein fragments and independently grouped into two populations by *k*-means clustering on the basis of their number of partners. Interactions involving fragments from groups with the greater number of partners were deemed promiscuous and removed from the final data set. Additional information is provided in Supplementary Methods.

Computational analysis. The identification of local regions of the protein interaction network with enhanced connectivity, comparisons of protein interaction data with *P. falciparum* microarray data sets from refs 16 and 17, and the discovery of proteins whose partners were enriched for protein domain or GO annotations were performed as described in Supplementary Methods.

Received 14 April; accepted 1 August 2005.

1. Breman, J. G. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**, 1–11 (2001).
2. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
3. Gardner, M. J. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
4. Sibley, C. H. et al. Yeast as a model system to study drugs effective against apicomplexan proteins. *Methods* **13**, 190–207 (1997).
5. Formstecher, E. et al. Protein interaction mapping: a *Drosophila* case study. *Genome Res.* **15**, 376–384 (2005).
6. Giot, L. et al. A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
7. Barabási, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
8. Rives, A. W. & Galitski, T. Modular organization of cellular networks. *Proc. Natl Acad. Sci. USA* **100**, 1128–1133 (2003).
9. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
10. Fan, Q., An, L. & Cui, L. *Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. *Eukaryot. Cell* **3**, 264–276 (2004).
11. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
12. Ragvin, A. et al. Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
13. Li, L., Stoekert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
14. Iida, T. & Araki, H. Noncompetitive counteractions of DNA polymerase epsilon and ISW2/yCHRC for epigenetic inheritance of telomere position effect in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24**, 217–227 (2004).
15. Huynh, K. D. & Bardwell, V. J. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **17**, 2473–2484 (1998).
16. Bozdech, Z. et al. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).
17. Le Roch, K. G. et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
18. Coulson, R. M., Hall, N. & Ouzounis, C. A. Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res.* **14**, 1548–1554 (2004).
19. Aravind, L., Iyer, L. M., Wellem, T. E. & Miller, L. H. *Plasmodium* biology: genomic gleanings. *Cell* **115**, 771–785 (2003).
20. Daniel, J. A. et al. Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. *J. Biol. Chem.* **279**, 1867–1871 (2004).
21. Collart, M. A. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* **313**, 1–16 (2003).
22. Ge, H., Walhout, A. J. & Vidal, M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* **19**, 551–560 (2003).
23. Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
24. Le Roch, K. G. et al. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res.* **14**, 2308–2318 (2004).

25. Li, X. *et al.* A co-ligand complex anchors *Plasmodium falciparum* merozoites to the erythrocyte invasion receptor band 3. *J. Biol. Chem.* **279**, 5765–5771 (2004).
26. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
27. Kissinger, J. C. *et al.* The *Plasmodium* genome database. *Nature* **419**, 490–492 (2002).
28. Cooke, B. M., Lingelbach, K., Bannister, L. H. & Tilley, L. Protein trafficking in *Plasmodium falciparum*-infected red blood cells. *Trends Parasitol.* **20**, 581–589 (2004).
29. Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
30. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Duffy, J. Feagin and C. H. Sibley for reading the manuscript critically, A. Gauntlett for technical assistance, and W. Hol for helpful discussions. This work was supported by a grant from the NIH. J.R.H. was supported by an NIH Kirschstein NRSA post-doctoral fellowship. S.F. is an Investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.F. (fields@u.washington.edu) or R.E.H. (rhughes@buckinstitute.org).

LETTERS

The *Plasmodium* protein network diverges from those of other eukaryotes

Silpa Suttham^{1,2*}, Taylor Sittler^{2*} & Trey Ideker^{1,2}

Plasmodium falciparum is the pathogen responsible for over 90% of human deaths from malaria¹. Therefore, it has been the focus of a considerable research initiative, involving the complete DNA sequencing of the genome², large-scale expression analyses^{3,4}, and protein characterization of its life-cycle stages⁵. The *Plasmodium* genome sequence is relatively distant from those of most other eukaryotes, with more than 60% of the 5,334 encoded proteins lacking any notable sequence similarity to other organisms². To systematically elucidate functional relationships among these proteins, a large two-hybrid study has recently mapped a network of 2,846 interactions involving 1,312 proteins within *Plasmodium*⁶. This network adds to a growing collection of available interaction maps for a number of different organisms, and raises questions about whether the divergence of *Plasmodium* at the sequence level is reflected in the configuration of its protein network. Here we examine the degree of conservation between the *Plasmodium* protein network and those of model organisms. Although we find 29 highly connected protein complexes specific to the network of the pathogen, we find very little conservation with complexes observed in other organisms (three in yeast, none in the others). Overall, the patterns of protein interaction in *Plasmodium*, like its genome sequence, set it apart from other species.

With the recent accumulation of protein interaction maps in public databases, cross-species comparisons are becoming critical for analysing the large networks formed by these interactions to delineate protein function and evolution⁷. At a fundamental level, protein networks can be compared to identify 'interologues'—that is, interactions that are conserved across species⁸. Beyond the individual comparison of interactions, methods such as PathBLAST (refs 9, 10) create a global alignment between protein networks to identify dense clusters of conserved interactions, suggestive of protein complexes. Such comparative approaches are important because they can tease conserved components of cellular machinery out of a highly connected network and increase overall confidence in the underlying interaction measurements.

We compared the protein–protein interaction network of *Plasmodium* reported by LaCount *et al.* (ref. 6) to protein networks for the budding yeast *Saccharomyces cerevisiae*¹¹, the nematode worm *Caenorhabditis elegans*¹², the fruitfly *Drosophila melanogaster*¹³ and the bacterial pathogen *Helicobacter pylori*¹⁴. Surprisingly, the pairwise alignment of these networks using PathBLAST (ref. 9) revealed that *Plasmodium* had only three conserved complexes with yeast (Fig. 1a–c), and no conserved complexes with any of the other organisms examined. However, yeast, fly and worm shared substantial numbers of conserved complexes with each other (Fig. 2a); for instance, yeast and fly had the highest degree of conservation with 61 conserved complexes.

The relatively low similarity between the *Plasmodium* protein

network and those of the other eukaryotes suggested that it encodes important functional differences worthy of further investigation. Alternatively, it was possible that differences in the number of complexes were related to network size. Thus, in addition to searching for conserved complexes, we investigated whether the observed similarities and differences were reflected in the probability of conservation of each protein interaction individually (Fig. 2b). For each pair of species, a protein–protein interaction was considered 'conserved' if both proteins had homologues that interacted in the opposite species (BLAST *E* value $\leq 1 \times 10^{-4}$, normalized for genome size). A global pairwise similarity metric was then defined as the overall fraction of interactions that were conserved, restricted to proteins with at least one homologue in the opposite species.

Figure 2c expresses the pairwise interaction similarities as a phylogenetic tree drawn using the method of Kitsch¹⁵. This tree was relatively robust to sampling errors as determined by bootstrap analysis: 86.2% of trials placed *Plasmodium* as an outgroup relative to yeast, worm and fly. Among the three model eukaryotes, yeast and worm were closest on the basis of interaction similarity (Fig. 2b), while yeast and fly were closest on the basis of conserved complexes (Fig. 2a). This discrepancy was probably due to network size or coverage. Nonetheless, the particular phylogenetic placement of *Plasmodium* was consistent across both analyses, and also agrees with the accepted taxonomical relationships among these organisms as established by morphological and sequence comparisons².

Another possibility for the low similarity of the *Plasmodium* protein network to the other species was that its interaction network had been measured predominantly among proteins expressed in the asexual stages of the parasite's life cycle (see ref. 6). There are two ways in which this sampling could have affected network similarity. First, it was possible that a high (or low) level of messenger RNA expression increases (or decreases) the number of interactions identified for the corresponding proteins, and thus alters the topology of the *Plasmodium* network relative to the other species. However, as shown in Supplementary Fig. 1, we found no correlation between expression level in any stage and the number of protein interactions. Second, it was possible that proteins from asexual stages tended to have lower similarity across species than proteins from other stages of the *Plasmodium* life cycle. However, we found that the *Plasmodium* interaction set was enriched for proteins with homologues in the other species, and that the protein interaction networks from all five organisms were enriched for yeast homologues in particular (Table 1; Supplementary Table 1). Such enrichment was observed even in worm, for which baits were explicitly selected to be non-homologous to yeast¹². This effect requires further study, but might indicate a bias of the yeast two-hybrid system in measuring interactions among yeast homologues, because all two-hybrid constructs must be expressible in the yeast cell.

A final possibility for the low level of conservation between the

¹Bioinformatics Program and ²Department of Bioengineering, University of California, San Diego, California 92093, USA.

*These authors contributed equally to this work.

Plasmodium protein network and those of the other organisms was that the *Plasmodium* network might have a substantially higher proportion of false-positive interactions relative to the networks of yeast, fly and worm. Lacking a 'gold standard' set of true interactions, we characterized the relative quality of the *Plasmodium* network by examining: (1) its global topological properties, and (2) the signal-to-noise ratio (SNR) of its protein complexes. Several common topological measures¹⁶ were computed on each network, including

the average number of interactions per protein (average degree), the average shortest path length between proteins, and the average clustering coefficient (Table 1). The number of interactions per protein in the *Plasmodium* interaction network followed a scale-free distribution, similar to other networks (Fig. 3a). Moreover, the *Plasmodium* network was never the outlier in any of the various measurements, suggesting that its global organization was consistent with the others.

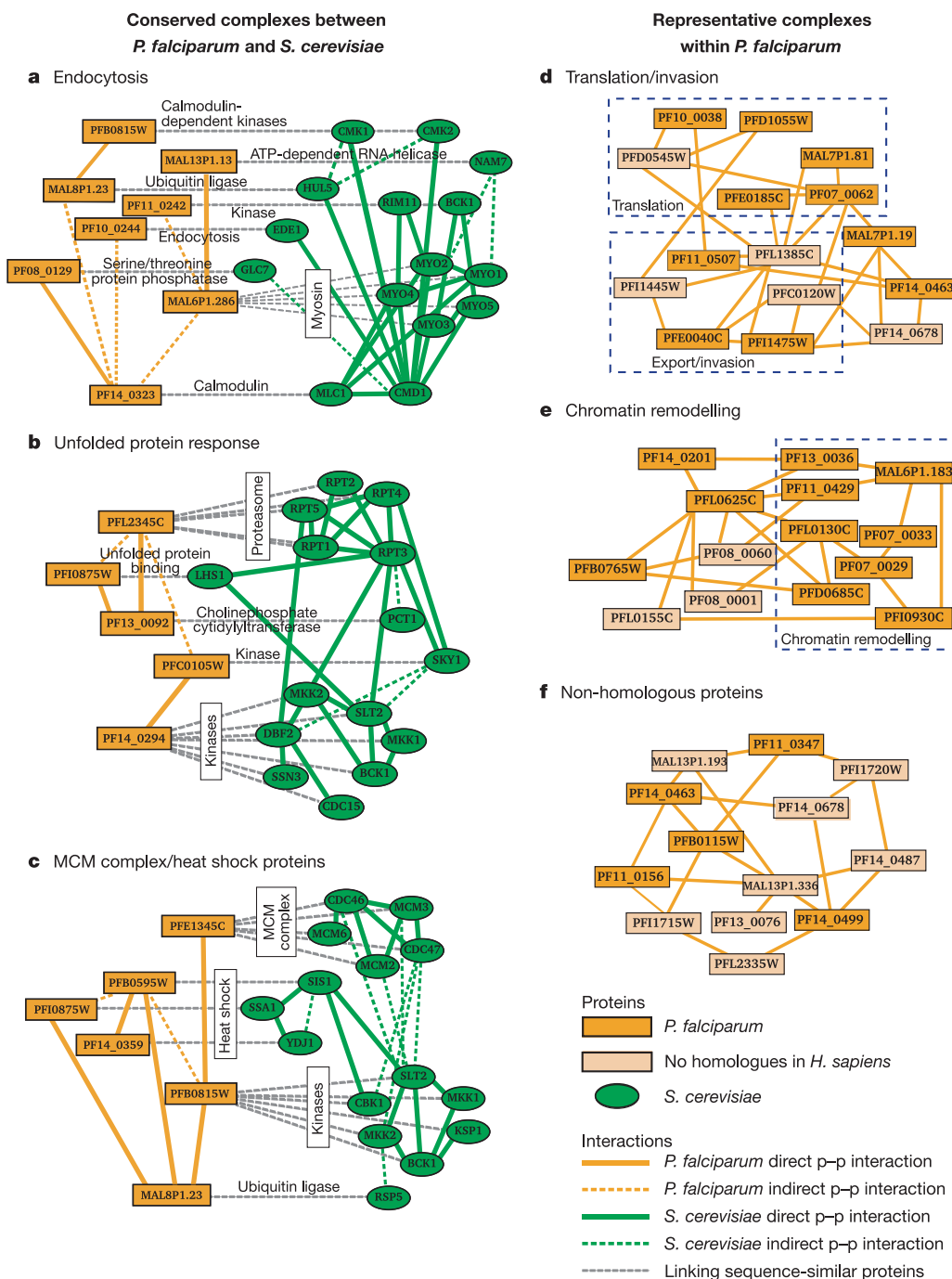


Figure 1 | Conserved and distinct complexes within *P. falciparum*. **a–c**, The three conserved complexes identified between *P. falciparum* and *S. cerevisiae*. Orange versus green nodes correspond to *P. falciparum* versus *S. cerevisiae* proteins. Solid links represent direct protein–protein (p–p) interactions, while dashed links represent indirect protein–protein interactions mediated by one other protein. Grey dashed lines connect sequence-similar proteins across the two species. MCM, minichromosome maintenance. **d–f**, Three

representative complexes found within the *P. falciparum* network only (out of 29 total complexes; see Supplementary Table 7). Cream-coloured nodes denote *Plasmodium* proteins without human homologues (*Homo sapiens*; using a permissive BLAST *E* value threshold $\leq 1 \times 10^{-2}$ to allow for distant homologues). All figures were drawn using the software Cytoscape (<http://www.cytoscape.org>).

Next, we applied the PathBLAST procedure to identify dense interaction complexes within each organism independently. A total of 29 single-species complexes were identified for *Plasmodium*, three of which are shown in Fig. 1d–f. This number was the median of the range observed over the five species (Table 1). Single-species complexes were used to assess the overall quality of each network by computing their SNR, a standard measure of assessing data quality in information theory and signal processing¹⁷. The SNR was computed by comparing the scores of complexes identified in the observed versus random interaction data for each organism (see the Methods). *Plasmodium*, worm and fly had very similar SNR values (Fig. 3b), while the SNR of the yeast network was slightly higher, and that of the

H. pylori network was slightly lower. The network distances of *Plasmodium* versus yeast, worm or fly do not appear to depend on SNR.

As the observed network differences could not be attributed to bias or error, we next examined the novel functional predictions suggested by the three conserved and 29 *Plasmodium*-specific complexes. The conserved protein complex shown in Fig. 1a predicts that the proteins PF10_0244 and MAL6P1.286 may have previously uncharacterized roles in endocytosis. The counterpart of PF10_0244 in the yeast network, Ede1, localizes to the cortical patch¹⁸ of the cell membrane at sites of polarized growth and seems to be involved in endocytosis¹⁹. Myo5 and Myo3, yeast

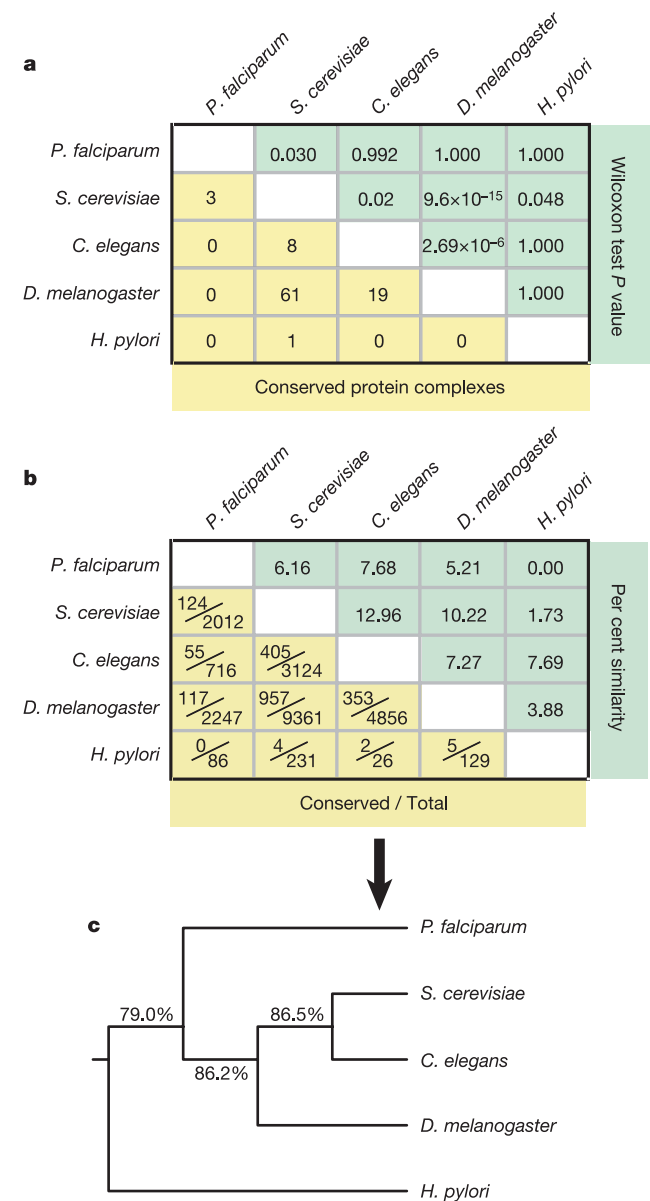


Figure 2 | Network similarity across the five organisms. **a**, The results of all pairwise PathBLAST comparisons. The number of conserved complexes is shown for each pair of species (yellow). The Wilcoxon rank-sum test *P* value (green) represents the significance of the distribution of all complex scores versus the distribution of complex scores found in equivalent random networks. **b**, The interaction-by-interaction similarity between networks is reported as both fractional values (yellow) and percentages (green). **c**, A phylogenetic tree constructed using these similarities. Percentages indicate the reproducibility of each branch during bootstrap analysis.

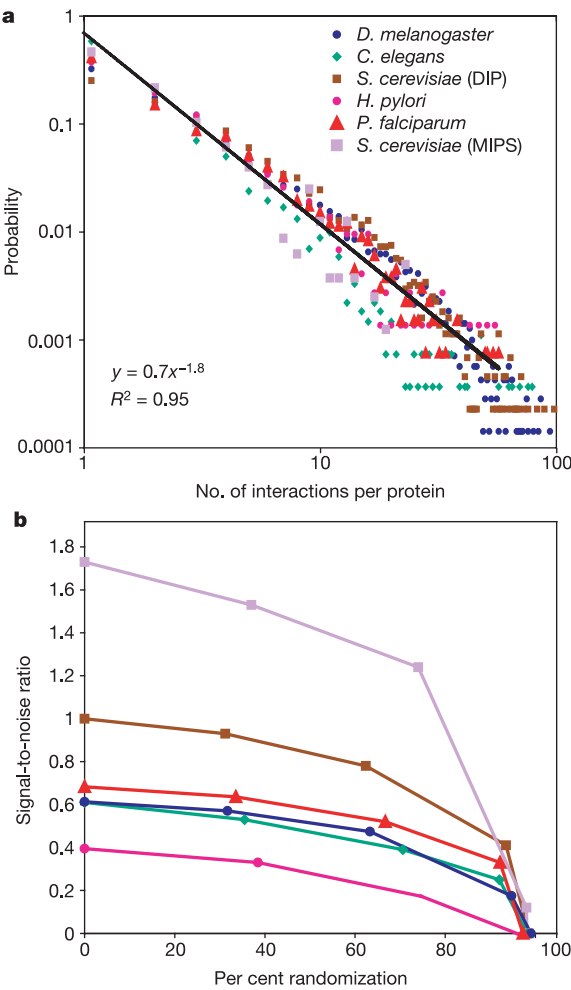


Figure 3 | Number of interactions per protein and the signal-to-noise ratio of protein complexes. **a**, Scale-free network behaviour is shown in a manner similar to ref. 16. The linear fit is for *Plasmodium* data only. **b**, Dependence of SNR on error rate. See the key in panel **a** for an explanation of the symbols. Each network was modified by randomly shuffling from 0–100% of its protein interactions to simulate the addition of false positives and negatives. The subsequent decrease in SNR, converging to SNR = 0 at 100% noise, validates that each network contains a substantial fraction of true positive interactions. The high-throughput *S. cerevisiae* data shown is from the network of yeast interactions in the Database of Interacting Proteins (DIP)¹¹. In addition to the high-throughput networks in this study, a literature-curated network—*S. cerevisiae* physical interactions according to the Munich Information Center for Protein Sequences (MIPS)³⁰—is provided as a positive control. Similar trends are observed for the average clustering coefficient and the per cent of interactions covered by established protein-domain interactions (Supplementary Table 3).

Table 1 | Topological properties of protein interaction networks

Organism	No. of interactions	Proteins covered	Average degree	Average shortest path	Average clustering coefficient*	No. of yeast homologues† (P value)	No. of single-species complexes
<i>S. cerevisiae</i> (DIP)‡	14,319	4,389	6.53	4.12	0.193	–	145
<i>S. cerevisiae</i> (Uetz)‡	1,449	1,345	2.16	6.95	0.049	–	66
<i>P. falciparum</i>	2,846	1,312	4.35	4.20	0.032	286 (2×10^{-133})	29
<i>D. melanogaster</i>	20,720	7,038	5.89	4.70	0.019	2,429 (2×10^{-205})	296
<i>C. elegans</i>	3,926	2,718	2.89	5.10	0.031	673 (4×10^{-10})	12
<i>H. pylori</i>	1,465	732	4.00	4.15	0.063	143 (1×10^{-2})	21

*The clustering coefficient measures local density of the network around a protein and is computed as previously described¹⁶.

†Yeast homologues are determined using a conservative BLAST *E* value threshold of $\leq 1 \times 10^{-10}$. The *P* values score the significance of enrichment for yeast homologues within the set of proteins covered by each interaction network, using the hypergeometric test. These enrichments are significant over a broad range of *E* value thresholds (data not shown).

‡Unlike other networks that are generated from single two-hybrid studies, the network of yeast interactions in the Database of Interacting Proteins (DIP)¹¹ consists of many experiments and experimental types. A separate analysis is included considering only the data from a single two-hybrid screen by Uetz *et al.* (ref. 29).

counterparts of MAL6P1.286, are class I myosins that also localize to actin cortical patches²⁰, where the calmodulin protein Cmd1 has been implicated in the uptake step of receptor-mediated endocytosis²¹. Taken together, this evidence suggests a role for this complex in calmodulin-mediated endocytosis. Calmodulin inhibitors have been shown to attenuate growth²² and chloroquine extrusion (thus effecting drug resistance)²³ in malarial parasites, and endocytosis has recently been linked to the mechanism of anti-malarial drugs including chloroquine and artemisinin²⁴. The proximity of calmodulin to the formation of endocytic vacuoles in *Plasmodium* provides for a discrete hypothesis linking endocytosis, drug resistance and drug mechanism-of-action.

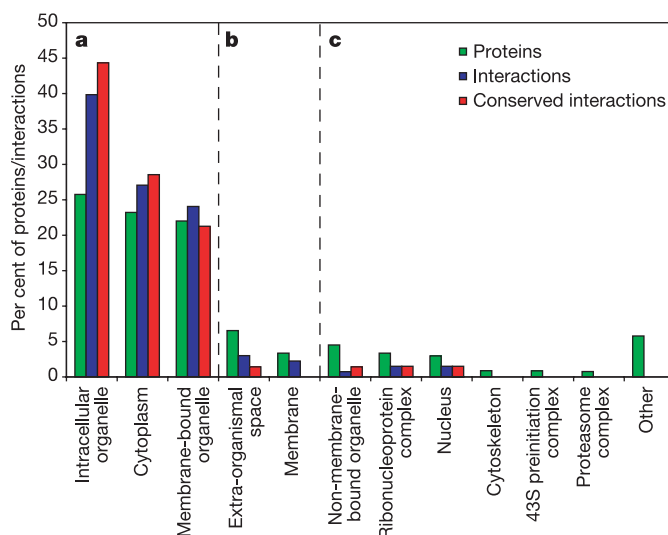
Within the 29 *Plasmodium*-specific complexes, chromatin remodelling was a prominent function, as shown in Fig. 1e. This complex involves the chromatin-remodelling protein ISWI (MAL6P1.183) interacting with a nucleosome assembly protein (PFI0930C)²⁵. The protein PF11_0429 has a PHD domain (for plant homeodomain), and PF07_0029 has an HSP90 domain (for heat shock protein of 90 kDa), both postulated to be involved in the remodelling process^{25,26}. Together, these known functions suggest that other proteins in the complex, such as PF08_0060, PFB0765W and PFL0625C, also participate in chromatin remodelling. For instance, although PFL0625C is annotated as a translation initiation factor, its yeast homologue has been found in complex with histone acetyltransferases²⁷. Further analysis of other complexes shown in Fig. 1 is available in the Supplementary Information.

Several cellular components that we expected to be present, such as the proteasome, were missing from the set of complexes conserved between *Plasmodium* and the other species. To investigate this issue, we plotted the distributions of known functional annotations (according to Gene Ontology Cellular Component Level Three)²⁸ among *Plasmodium* proteins, protein interactions and conserved interactions (Fig. 4 and Supplementary Fig. 2; note that a protein or an interaction can participate in multiple categories). Considerable proportions of all three data sets were associated with intracellular organelles, membrane-bound organelles or the cytoplasm (Fig. 4a). Other cellular components, such as the membrane and extra-organismal space, were represented among proteins and interactions but to a lesser extent among conserved interactions (Fig. 4b). Many membrane-associated components were also reported in the 29 *Plasmodium*-specific complexes, and are suggestive of machinery unique to this organism. Finally, components such as the proteasome and cytoskeleton were represented among proteins but were absent from the interaction set, and hence were not found as conserved interactions or complexes (Fig. 4c). Interactions among proteins in these components may have yet to be uncovered. These observations are reinforced by a complementary analysis of the functional distributions of yeast, worm and fly protein networks (Supplementary Fig. 2).

In summary, we have characterized conserved patterns of interaction between the protein network of *Plasmodium falciparum* and those of other species, and reported the specific network regions that

are conserved. All of the examined networks contain dense complex-like structures of interactions, some of which are shared by yeast, worm and fly but not *Plasmodium*. These relationships are not clearly related to noise or bias in the *Plasmodium* interaction set. Some of the observed differences are almost certainly due to incomplete coverage in one or more networks: for instance, the present *Plasmodium* interaction set is focused on asexual life-cycle stages. Nevertheless, our comparison reflects the relative degree of similarity between the different networks. These differences are observed even when considering only those genes that are homologous across species.

It is generally expected that conserved genes will retain their functions and interactions. From this comparison, a different principle emerges: conservation of specific groups of related genes does not necessarily imply conservation of interaction among their encoded proteins. Further studies may distinguish the true differences from those related to network coverage and, ultimately,

**Figure 4 | Functional roles within the *Plasmodium* protein network.**

Functional annotations within the *Plasmodium* protein network associated with intracellular organelles, membrane-bound organelles or the cytoplasm (a), the membrane and extra-organismal space (b), and other components such as the proteasome and cytoskeleton (c). The histograms show the distribution of Gene Ontology Cellular Component assignments over all annotated *Plasmodium* proteins (green bars), protein interactions (blue bars) and conserved protein interactions (red bars). Interactions are considered 'annotated' if the interacting proteins share the same Gene Ontology category (these interactions are listed in Supplementary Table 6). For conserved interactions, the percentages in each category are cumulative over the three pairwise comparisons of *Plasmodium* versus the other three eukaryotes (yeast, fly or worm). Note that a protein or interaction can participate in multiple categories.

facilitate the discovery of new pharmaceuticals directed at the protein complexes unique to this parasite.

METHODS

Identification of conserved and species-specific complexes. Identification of protein complexes was performed using the PathBLAST family of network alignment tools, as previously described⁹. Briefly, these methods integrate protein interaction data from two species with protein sequence homology to generate an 'aligned network', in which each node represents a pair of homologous proteins (one from each organism; BLAST *E* value $\leq 1 \times 10^{-10}$) and each link represents a conserved interaction. The network alignment is searched to identify high-scoring subnetworks, for which the score is based on the density of interactions within the subnetwork as well as confidence estimates for each protein interaction (see below). The search is then repeated over 100 random trials, in which the interactions of both species are arbitrarily reassigned while maintaining the same number of interactions per protein, resulting in a distribution of random subnetwork scores pooled over all trials. Dense subnetworks that score in the top fifth percentile of this random score distribution are considered significant and reported as 'conserved complexes'. The search for 'single-species complexes' is identical to the search for conserved complexes, except that an individual protein network is searched instead of the network alignment. This process identifies dense subnetworks constrained by the interactions of one organism rather than two.

Interaction confidence scores. We estimated the probability that each measured protein interaction is true using a logistic regression model based on mRNA expression correlation, the network cluster coefficient, and the number of times the interaction had been experimentally observed. Further information on these confidence assignments is provided in the Supplementary Methods.

Phylogenetic tree construction. The Kitsch algorithm (provided by the PHYLIP package¹⁵) assumes the presence of an evolutionary clock and is based on pairwise distances between species. For each pair of species, an interaction between proteins *a* and *b* was considered 'conserved' if both proteins had sequence-similar counterparts *a'* and *b'* (BLAST *E* value $\leq 1 \times 10^{-4}$) that interacted in the opposite species. A pairwise similarity between networks was computed as $s_{1,2} = (c_1 + c_2)/(t_1 + t_2)$, where *c* is the number of conserved interactions and *t* is the total number of interactions in species 1 or 2, respectively (with all interactions restricted to the set of proteins with homologues in the opposite species). Pairwise network distance was then defined as $1 - s_{1,2}$. The resulting phylogenetic tree shown in Fig. 2c is the consensus over 10,000 bootstrap simulations. Values of *c* and *t* for each network are listed in Supplementary Table 2.

Signal-to-noise ratio of protein complexes. SNR was computed for the single-species complexes as follows. The search for dense interaction complexes is initiated from each node (protein) and the highest scoring complex from each is reported (see the 'Identification of conserved and species-specific complexes' section of the Methods). This yields a distribution of complex scores over all nodes in the network. A score distribution is also generated for 100 randomized networks, which have an identical degree distribution as the original network. The SNR ratio is computed from these original and random score distributions (representing signal and noise, respectively) according to the standard formula¹⁷ using the root mean square (r.m.s.):

$$\text{SNR} = \log_{10} \frac{\text{r.m.s. (original complex scores)}}{\text{r.m.s. (random complex scores)}}, \quad (1)$$

$$\text{with r.m.s. (x)} = \sqrt{\frac{1}{M} \sum_{i=1}^M x_i^2}$$

and where x_i is the score of a complex and *M* is the total number of complexes.

Received 22 April; accepted 8 August 2005.

1. Miller, L. H., Baruch, D. I., Marsh, K. & Doumbo, O. K. The pathogenic basis of malaria. *Nature* **415**, 673–679 (2002).
2. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
3. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).
4. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
5. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
6. LaCount, D. J. *et al.* A protein interaction network of the malaria parasite

Plasmodium falciparum. *Nature* doi:10.1038/nature04104 (this issue).

7. Conant, G. C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genet.* **34**, 264–266 (2003).
8. Yu, H. *et al.* Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* **14**, 1107–1118 (2004).
9. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. USA* **102**, 1974–1979 (2005).
10. Kelley, B. P. *et al.* Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl Acad. Sci. USA* **100**, 11394–11399 (2003).
11. Xenarios, I. *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**, 303–305 (2002).
12. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
13. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
14. Rain, J. C. *et al.* The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–215 (2001).
15. Felsenstein, J. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166 (1989).
16. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
17. Shanmugam, K. S. *Digital and Analog Communication Systems* (Wiley, New York, 1979).
18. Gagny, B. *et al.* A novel EH domain protein of *Saccharomyces cerevisiae*, Ede1p, involved in endocytosis. *J. Cell Sci.* **113**, 3309–3319 (2000).
19. Engqvist-Goldstein, A. E. & Drubin, D. G. Actin assembly and endocytosis: from yeast to mammals. *Annu. Rev. Cell Dev. Biol.* **19**, 287–332 (2003).
20. Goodson, H. V., Anderson, B. L., Warrick, H. M., Pon, L. A. & Spudich, J. A. Synthetic lethality screen identifies a novel yeast myosin I gene (MYO5): myosin I proteins are required for polarization of the actin cytoskeleton. *J. Cell Biol.* **133**, 1277–1291 (1996).
21. Salisbury, J. L., Condeelis, J. S., Maihle, N. J. & Satir, P. Calmodulin localization during capping and receptor-mediated endocytosis. *Nature* **294**, 163–166 (1981).
22. Scheibel, L. W. *et al.* Calcium and calmodulin antagonists inhibit human malaria parasites (*Plasmodium falciparum*): implications for drug design. *Proc. Natl Acad. Sci. USA* **84**, 7310–7314 (1987).
23. Sanchez, C. P., McLean, J. E., Stein, W. & Lanzer, M. Evidence for a substrate specific and inhibitable drug efflux system in chloroquine resistant *Plasmodium falciparum* strains. *Biochemistry* **43**, 16365–16373 (2004).
24. Hoppe, H. C. *et al.* Antimalarial quinolones and artemisinin inhibit endocytosis in *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **48**, 2370–2378 (2004).
25. Langst, G. & Becker, P. B. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* **114**, 2561–2568 (2001).
26. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33**, 70–74 (2003).
27. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
28. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
29. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
30. Mewes, H. W. *et al.* MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to M. Vignali, D. LaCount and S. Fields at the University of Washington, and B. Hughes and S. Sahasrabudhe at Prolexys, for providing us with advance access to the *Plasmodium* interaction data and for suggestions on our manuscript. We also thank E. Winzeler and J. Vinetz for advice on *Plasmodium* protein function, R. Sharan for help with the PathBLAST algorithm, and V. Bafna for assistance with the false-positive analysis. Finally, we acknowledge the following funding support: the National Science Foundation (S.S.); the National Institute of General Medical Sciences (T.I.); a David and Lucille Packard Fellowship award (T.I.); the Howard Hughes Medical Institute (T.S.); and Unilever (T.S.).

Author Contributions S.S. and T.S. contributed equally to this work. All authors discussed the results and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.S. (ssuthram@ucsd.edu).

A putative stimulatory role for activator turnover in gene expression

J. Russell Lipford¹, Geoffrey T. Smith¹, Yong Chi^{1†} & Raymond J. Deshaies¹

The ubiquitin–proteasome system (UPS) promotes the destruction of target proteins by attaching to them a ubiquitin chain that is recognized by the 26S proteasome¹. The UPS influences most cellular processes, and its targets include transcriptional activators that are primary determinants of gene expression. Emerging evidence indicates that non-proteolytic functions of the UPS might stimulate transcriptional activity^{2,3}. Here we show that the proteolysis of some transcriptional activators by the UPS can stimulate their function. We focused on the role of UPS-dependent proteolysis in the function of inducible transcriptional activators in yeast, and found that inhibition of the proteasome⁴ reduced transcription of the targets of the activators Gcn4, Gal4 and Ino2/4. In addition, mutations in SCF^{Cdc4}, the ubiquitin ligase for Gcn4 (ref. 5), or mutations in ubiquitin that prevent degradation⁶, also impaired the transcription of Gcn4 targets. These transcriptional defects were manifested despite the enhanced abundance of Gcn4 on cognate promoters. Proteasome inhibition also decreased the association of RNA polymerase II with Gcn4, Gal4 and Ino2/4 targets, as did mutations in SCF^{Cdc4} for Gcn4 targets. Expression of a stable phospho-site mutant of Gcn4 (ref. 7) or disruption of the kinases that target Gcn4 for turnover^{5,7} alleviated the sensitivity of Gcn4 activity to defects in the UPS.

The UPS is a fundamental component of normal cell growth and proliferation. The UPS is also important for cancer cell growth, as highlighted by the recent approval of the proteasome inhibitor, Velcade, for the treatment of relapsed multiple myeloma⁸. Recent studies have investigated the mechanism of Velcade action by examining the transcriptional response to the drug in human and yeast cells^{9,10}. These studies indicate that proteasome inhibition does not substantially alter bulk transcription of the genome in myeloma cells or in *Saccharomyces cerevisiae*. However, a group of genes are repressed by Velcade, including human growth and survival genes and yeast genes involved in the biosynthesis of amino acids^{9,10}. These yeast genes are regulated by the b-ZIP transcriptional activator Gcn4, which promotes the expression of more than 500 genes¹¹. Gcn4 is a target for UPS-mediated degradation¹² through the E3-ubiquitin ligase SCF^{Cdc4}. Ligases comprise the final, substrate recognition step of the ubiquitination cascade. SCF^{Cdc4} ubiquitinates and targets for proteolysis Gcn4 molecules that have been phosphorylated by the cyclin-dependent kinases (CDKs) Srb10 and Pho85 (refs 5, 7). We have attempted to explain the role of the UPS in the function of Gcn4 and other activators.

To assess the impact of proteolysis on Gcn4 function, we treated a yeast strain (*pdr5Δ*) that is sensitive to proteasome inhibitors¹⁰ with the Velcade analogue MG132 (ref. 4). Reverse transcriptase-mediated polymerase chain reaction (RT–PCR) analyses confirmed that, as for Velcade¹⁰, treatment with MG132 substantially reduced the transcription of several Gcn4 targets, including *HIS4* and *CPA2*, in

minimal medium, in comparison with dimethylsulphoxide (DMSO) alone (Fig. 1a and Supplementary Fig. S1a). Similar results were obtained when Gcn4 was highly induced by amino acid starvation (Fig. 1a, –Leu; see Supplementary Information for discussion of media). These effects were largely dependent on Gcn4, because we observed similar decreases in the transcription of a reporter driven exclusively by six Gcn4-binding sites¹³ (Fig. 1a, *GCRE6–LacZ*).

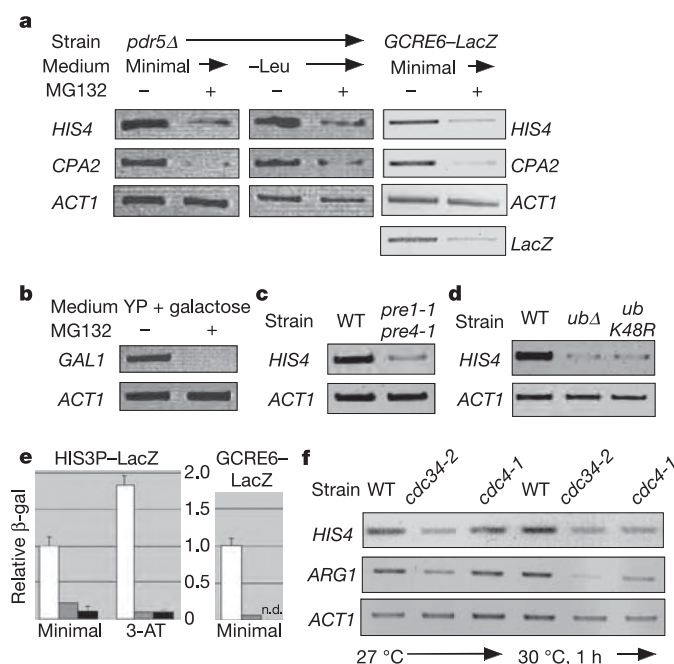


Figure 1 | UPS-dependent proteolysis positively regulates inducible transcriptional activators. **a**, Indicated strains were grown in minimal or starvation (–Leu) medium, treated with MG132 (50 μM) or DMSO, and processed for RT–PCR of the indicated transcripts. *pdr5Δ* enables the uptake of MG132 into yeast. **b**, A *pdr5Δ* strain was induced with galactose, treated with MG132 and prepared for RT–PCR of *GAL1* and *ACT1*. **c**, wild-type (WT) and *pre1-1, pre4-1* strains were grown in minimal medium at 27 °C and processed for RT–PCR of *HIS4* and *ACT1*. **d**, Conditionally expressed ubiquitin was depleted from strains while expression from a complementing plasmid encoding either WT ubiquitin, no ubiquitin (Δ) or K48R-ubiquitin was induced (see Supplementary Information). Samples were prepared for RT–PCR as above. **e**, WT (white bars), *cdc34-2* (grey bars) and *gcn4Δ* (black bars) strains expressing LacZ from the *HIS3* or *GCRE6* promoter were grown in minimal or starvation medium (3-AT) at 30 °C and then processed to test β-galactosidase (β-gal) activity. Standard deviations are from three replicates. n.d., not detectable. **f**, WT, *cdc34-2* and *cdc4-1* strains were grown at 27 °C or shifted to 30 °C for 1 h. RT–PCR was performed for *HIS4*, *ARG1* and *ACT1*.

¹Howard Hughes Medical Institute, Division of Biology, MC 156-29, California Institute of Technology, 1200 E. California Boulevard, Pasadena, California 91125, USA.

[†]Present address: Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA.

We also tested the proteasome dependence of two regulons (GAL and INO) that might have eluded the Velcade microarray analysis because they were not expressed under the growth conditions that were employed. MG132 substantially decreased induction of the Gal4 target, *GAL1*, on the addition of galactose (Fig. 1b). Similarly, on the removal of inositol, transcription of *INO1*, a target of the activators Ino2 and Ino4, was largely abrogated by MG132 (Supplementary Fig. S1b). In all cases treatment with MG132 did not affect the constitutive transcription of *ACT1*.

These findings indicate that the 26S proteasome might promote the function of some inducible transcriptional activators. Focusing on Gcn4, we tested additional components of the UPS for their impact on transcription. A strain (*pre1-1* and *pre4-1*) mutated for the peptidase activity of the proteasome showed reduced transcription of *HIS4* in comparison with a wild-type (WT) strain (Fig. 1c)¹⁴. We also tested transcription of *HIS4* in strains that conditionally express alternative versions of ubiquitin⁶. When endogenous ubiquitin was depleted in the presence of a vector plasmid (Ub Δ) or depleted in a cell expressing K48R ubiquitin, *HIS4* messenger RNA levels were sharply decreased in comparison with a depleted strain that expressed WT ubiquitin (Fig. 1d). Again, the *ACT1* mRNA remained constant. These findings confirm that ubiquitination and proteolysis are important for Gcn4 function. Importantly, because the K48R mutant cannot form chains that target substrates to the proteasome, these findings also indicate that, in contrast to the regulation proposed for Gal4–VP16 (ref. 15) and c-Myc^{16,17}, mono-ubiquitination might not be able to sustain Gcn4 activity.

We next examined the impact of Cdc34–SCF^{Cdc4}, the specific E2–E3 ubiquitin ligase for Gcn4 (refs 5, 7), on activator function. Gcn4-dependent expression of β -galactosidase was evaluated in WT, temperature-sensitive *cdc34-2*, and *gcn4 Δ* strains. After growth at the semi-permissive temperature of 30 °C, the *cdc34-2* strain exhibited a fourfold decrease (relative to WT) in reporter expressed from the *HIS3* promoter (*HIS3P*) in minimal medium and a roughly 15-fold reduction on starvation by 3-aminotriazole (3-AT, ref. 11) (Fig. 1e). Expression from GCRC6–LacZ was also compromised about 10-fold in *cdc34-2* (Fig. 1e). In addition, expression from both promoters was extinguished in *gcn4 Δ* cells under all conditions (Fig. 1e). RT–PCR analysis was then used to assess effects on endogenous targets. Gcn4-dependent transcription of *HIS4* and *ARG1* was defective in both *cdc34-2* and *cdc4-1* (a thermosensitive allele of the SCF F-box protein Cdc4) strains at 30 °C (Fig. 1f). In contrast, *ACT1* transcript levels in all strains were similar. These results point to a stimulatory role for Cdc34–SCF^{Cdc4} in Gcn4-mediated transcription.

To explore the molecular basis of the stimulation of Gcn4 function by the UPS, we examined the effect that proteasome inhibition has on the abundance and ubiquitination of Gcn4. Western blotting showed that MG132 increased the abundance of chromosomally encoded Gcn4–Myc9 and led to the appearance of a high-molecular-mass ladder (Fig. 2a, lanes 1 and 2). Immunoprecipitation and western analysis with anti-ubiquitin antibodies confirmed that this ladder was ubiquitinated Gcn4 (Fig. 2a, lanes 5, 6, 9 and 10). Because transcription mediated by Gcn4 was strongly repressed by MG132 (Fig. 1), ubiquitination was presumably insufficient to sustain Gcn4 activity. To confirm the specificity of Gcn4 ubiquitination, we repeated the analysis with a *gcn4-3T2S* strain. Gcn4–3T2S lacks five phosphorylation sites, is no longer phosphorylated by Srb10 or Pho85 nor ubiquitinated by SCF^{Cdc4} *in vitro*, and is stabilized⁷. As predicted, the 3T2S strain had much lower levels of ubiquitinated Gcn4 (Fig. 2a, lanes 3, 4, 7, 8, 11 and 12).

We next investigated promoter occupancy by Gcn4 and Gal4 and recruitment of RNA polymerase II (polII) to target genes. Chromatin immunoprecipitation (ChIP) assays¹⁸ revealed a substantial increase in the association of Gcn4–Myc9 with the *HIS4* promoter in cells treated with MG132 (Fig. 2b, 9E10). Similar experiments showed little change in levels of TAP-tagged Gal4 at the *GAL1* promoter on treatment with MG132 (Fig. 2c, IgG). The galactose-dependent

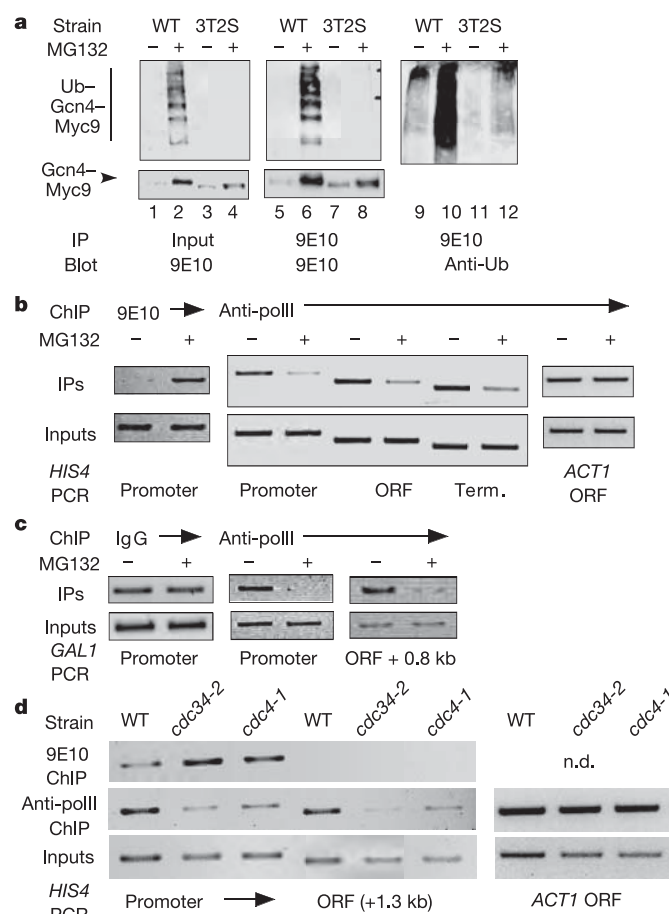


Figure 2 | Defects in the UPS lead to the accumulation of ubiquitinated Gcn4 and impair the association of RNA polymerase II with Gcn4 and Gal4 targets. **a**, *gcn4-3T2S* strains expressing Gcn4–Myc9 (WT) or Gcn4–3T2S–Myc9 (3T2S) were grown in minimal medium and treated with MG132 or DMSO. Immunoprecipitations (IPs) were performed with 9E10 antibodies recognizing Myc9, and subsequent western blots of the input (lanes 1–4) or immunoprecipitation (IP; lanes 5–12) samples were probed with 9E10 (lanes 1–8) or anti-Ub antibodies (lanes 9–12). **b**, A WT strain, as above, was processed for ChIP analysis with 9E10 and anti-polII antibodies. PCR was performed to amplify the promoter, ORF and terminator regions of *HIS4* and the ORF of *ACT1*. **c**, ChIP analysis of Gal4–TAP and polII was performed with galactose-induced strains treated with MG132 or DMSO. The promoter and ORF regions of *GAL1* were amplified. IgG was used to retrieve Gal4–TAP. **d**, ChIP analysis of Gcn4–Myc9 and polII was performed, as in **b**, with WT, *cdc34-2* and *cdc4-1* strains that were grown at 27 °C, then shifted to 30 °C for 1 h. kb, kilobases; n.d., not done.

decrease in promoter-bound Gal80–TAP was also unaffected by MG132 (Supplementary Fig. S3). When the ChIP was performed with polII, MG132 reduced the signal for the promoter, open reading frame (ORF) and terminator regions of the Gcn4 target *HIS4*, whereas the signal for the *ACT1* ORF was unaffected (Fig. 2b, polII). MG132 also decreased the polII ChIP signal for the *GAL1* promoter and ORF (Fig. 2c, polII) and the *INO1* promoter (data not shown).

The ChIP analysis was extended to *cdc34-2* and *cdc4-1* strains. At 30 °C, more Gcn4–Myc9 (Fig. 2d, 9E10) but less polII was associated with the *HIS4* promoter in the SCF mutants than in the WT. Recruitment of polII to *ACT1* was unaffected in the SCF mutants. Results from all the ChIP analyses closely parallel the transcription results and imply that the UPS is important for sustaining the interaction of RNA polymerase II with the targets of some activators despite the increased accumulation of activator (for Gcn4) at the promoter.

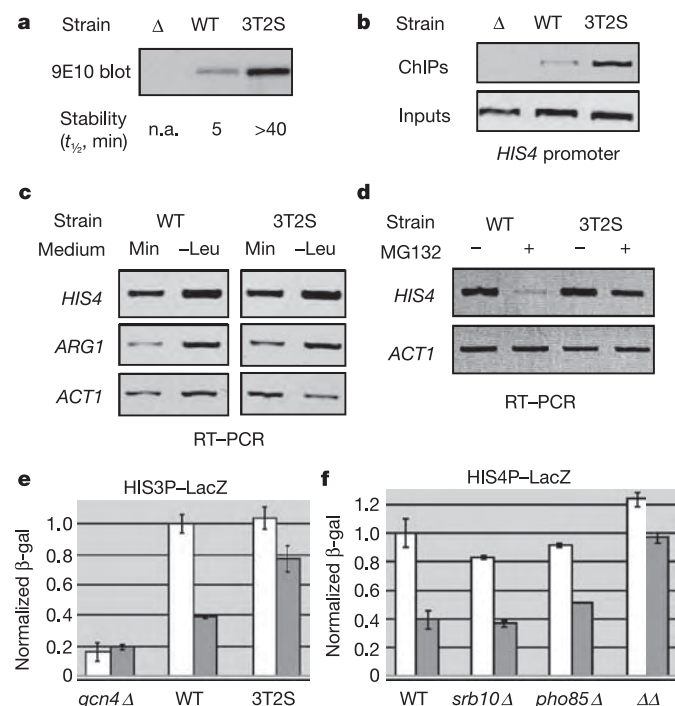


Figure 3 | The UPS has little effect on the activity of stable, non-phosphorylated versions of Gcn4. **a**, GCN4 (WT), *gcn4-3T2S* or *gcn4 Δ* strains were grown in minimal medium and processed for western blotting with 9E10 antibodies. Protein half-life data ($t_{1/2}$) were reported previously⁷. n.a., not applicable. **b**, ChIP analysis of the above strains was performed for the *HIS4* promoter. **c**, RT-PCR of the indicated transcripts were performed with WT and *gcn4-3T2S* strains grown in minimal (Min) or starvation (-Leu) medium. **d**, RT-PCR analysis of *HIS4* and *ACT1* was performed with WT and 3T2S strains in the presence and absence of MG132. **e**, *CDC34* (open bars) and *cdc34-2* (filled bars) strains with the *HIS3P-LacZ* reporter and expressing either WT, 3T2S or null (*gcn4 Δ*) versions of GCN4 were grown in minimal medium and processed for β -galactosidase (β -gal) activity. Standard deviations were calculated from three replicates. **f**, WT and *cdc34-2* strains with a *HIS4P-LacZ* reporter and harbouring WT or deleted versions of *SRB10* and/or *PHO85* were treated as in **e**.

To test whether turnover of the activator itself can promote function, we evaluated the stabilized *Gcn4-3T2S* (ref. 7). In minimal medium the levels of total protein (Fig. 3a) and promoter-associated *Gcn4-Myc9* (Fig. 3b) were about twofold to threefold higher in the *gcn4-3T2S* strain. Despite such increases, *gcn4-3T2S* did not alter the expression of *Gcn4* targets in minimal (Fig. 3c, SM +; Fig. 3e, *CDC34*) or starvation medium (Fig. 3c, -Leu), indicating a possible decrease in specific activity. Most importantly, *gcn4-3T2S* partly alleviated the deleterious effects of an impaired UPS on *Gcn4*-dependent transcription. For example, MG132 inhibited *HIS4* transcription in *gcn4-3T2S* much less than in *GCN4* (Fig. 3d). In addition, *cdc34-2* diminished *HIS3P-LacZ* expression only about 1.2-fold in *gcn4-3T2S*, in comparison with more than 2.5-fold in *GCN4* (Fig. 3e). The *cdc34-2* mutation had no impact on *gcn4 Δ* (Fig. 3e). A similar epistatic relationship was seen on deletion of *SRB10* and *PHO85*, which stabilizes *Gcn4* to a similar extent to the *gcn4-3T2S* mutation⁷. Deletion of both CDKs ($\Delta\Delta$) slightly increased *LacZ* expression (1.2-fold) and, as with *gcn4-3T2S*, *cdc34-2* only mildly affected expression (1.2-fold) in the $\Delta\Delta$ strain (Fig. 3f). The suppression of *cdc34-2* required the deletion of both kinases, because *srb10 Δ* and *pho85 Δ* single mutants remained relatively sensitive to *cdc34-2* (Fig. 3f). We note that *gcn4-3T2S* is not completely refractory to UPS inhibition, indicating that, in addition to *Gcn4*, the UPS might also promote transcription through other factors. Nevertheless, these findings indicate that proteolysis of CDK-phosphorylated *Gcn4* by means of the UPS might be important in sustaining

maximal expression of *Gcn4* targets and that, in the absence of phosphorylation, *Gcn4* activity is less dependent on its turnover.

Components of the UPS have been posited to activate transcription by multiple mechanisms^{2,3}. In numerous examples, including the activation of *Gcn4* in response to ultraviolet radiation¹⁹ and transcription of NF- κ B (ref. 2) and oestrogen receptor targets²⁰, the UPS seems to mediate signalling upstream of the activator. As discussed in Supplementary Fig. S3, this mechanism probably does not account for our findings. Meanwhile, subunits of the 19S cap of the proteasome have been suggested to have a positive role in transcription that is independent of their proteolytic function^{21,22}. In addition, it has been proposed that the ubiquitination of Gal4-VP16 and c-Myc transiently increases the activity of these factors before proteolysis¹⁵⁻¹⁷. Our results differ substantively from these examples, in that neither the 19S cap (whose activity is not known to be affected by inhibition of the 20S proteases²³) nor ubiquitination was sufficient to achieve maximal transcription of *Gcn4* targets (Figs 1d and 2a). Instead, we found that inhibition of the proteasome and genetic manipulations of the UPS, the CDKs for *Gcn4*, and *Gcn4* itself all provided evidence that turnover of *Gcn4* normally enhances its function.

Proteasome activity also seems to sustain inducible transcription mediated by Gal4 and Ino2/4. Interestingly, activation of promoter-associated Gal4 in galactose medium requires *Srb10*-dependent phosphorylation^{24,25}, and this activated form has a short half-life²⁴. This indicates that Gal4 might be regulated by degradation in a manner similar to that of *Gcn4*. We previously proposed a model consistent with these current findings in which proteolysis is required to remove 'spent' activators and to reset the promoter³. The initial 'pioneer round(s)' of transcription would not involve the UPS, but subsequent rounds would be stimulated by turnover of the spent, promoter-bound activator to allow binding of a fresh molecule. This mechanism places *Gcn4*, Gal4 and Ino2/4 into a class of regulatory factors—including securin, p21 and p27—whose activity is required early in a process but whose subsequent turnover or removal promotes completion of the process or subsequent reaction cycles. We call this phenomenon 'activation by destruction' and believe that, given the diversity of the examples listed in Supplementary Table S2, it might represent a regulatory mechanism for a large class of factors and might be an important determinant of infection and disease.

METHODS

Yeast strains, growth conditions and extract preparation. A complete list of yeast strains used in this study is provided in Supplementary Table S1. All strains were derived from the S288C background, except RJD 2505 and RJD 3137-3141, which were derived from the W303 background. Strains were constructed and grown in accordance with standard protocols²⁶. A description of the various media used in the study is given in Supplementary Information. MG132 (American Peptide) was added to cultures of *prf5 Δ* strains to a final concentration of 50 μ M. All extracts were prepared by lysis with glass beads (Sigma) in a Fast Prep (Bio 101) device. Ubiquitin derivative analysis (Fig. 1d) is described in the Supplementary Methods.

RT-PCR analysis. mRNA was prepared using RNeasy kits (Qiagen). Total mRNA (200 ng) and 10 pmol of oligo(dT) were used to reverse-transcribe complementary DNA (Stratagene). One-tenth of the cDNA reaction was then used for 20-22 cycles of PCR and products were resolved on 2% agarose gels. Primer sequences are available from the authors on request.

β -Galactosidase assays. The *HIS3P-LacZ*, *HIS4P-LacZ* and *GCRC-LacZ* reporter constructs and the protocol to measure β -galactosidase activity were as described previously²⁷. Reported activity was normalized to total extract protein as measured by bicinchoninic acid assay (Pierce). Relative activities are reported with average WT activity set to 1.

Western blots. Except as noted, extracts were prepared by immediate boiling of cell pellets in 2 $\times$ Laemmli SDS sample buffer followed by lysis with glass beads. Equal amounts of total protein (as judged by Coomassie staining) were resolved by SDS-PAGE. Blots were probed with 9E10 antibodies to recognize *Gcn4-Myc9* or FK1 antibodies (Affiniti) to recognize ubiquitin. Horseradish peroxidase-coupled goat anti-mouse secondary antibodies (Bio-Rad) were used for detection.

Immunoprecipitations and detection of ubiquitinated Gcn4. DMSO-treated or MG132-treated cultures were treated with formaldehyde (final concentration 1%) for 20 min to trap ubiquitinated intermediates. Crosslinking was quenched and extracts were made in ChIP buffer¹⁸. A 10% sample of the extract (whole cell extract) was removed and boiled in SDS sample buffer. The remainder of the extract was incubated at 4 °C with 9E10 antibodies coupled to Protein A–Sephadex (Sigma) beads. Proteins were eluted and crosslinks were reversed by being boiled in SDS sample buffer. Protein samples were then processed for western blotting. For FK1 (anti-ubiquitin) western blots, the nitrocellulose was boiled before incubation with the antibody.

ChIP assays. ChIP assays were performed as described¹⁸; 9E10 antibodies were used to immunoprecipitate chromatin fragments associated with Gcn4–Myc9 and antibodies against the carboxy-terminal domain of the largest subunit of RNA polymerase II (anti-polII; Covance) were used to immunoprecipitate fragments associated with polII. Protein G–Sephadex beads (Amersham Biosciences) were used to precipitate antibody–antigen complexes. Rabbit immunoglobulin (Ig) G–agarose (Sigma) was used to immunoprecipitate Gal4-associated fragments. Primer sequences are available from the authors on request.

Received 2 June; accepted 3 August 2005.

- Pickart, C. M. Back to the future with ubiquitin. *Cell* **116**, 181–190 (2004).
- Muratani, M. & Tansey, W. P. How the ubiquitin–proteasome system controls transcription. *Nature Rev. Mol. Cell Biol.* **4**, 192–201 (2003).
- Lipford, J. R. & Deshaies, R. J. Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation. *Nature Cell Biol.* **5**, 845–850 (2003).
- Lee, D. H. & Goldberg, A. L. Proteasome inhibitors: valuable new tools for cell biologists. *Trends Cell Biol.* **8**, 397–403 (1998).
- Meimoun, A. *et al.* Degradation of the transcription factor Gcn4 requires the kinase Pho85 and the SCF(CDC4) ubiquitin–ligase complex. *Mol. Biol. Cell* **11**, 915–927 (2000).
- Finley, D. *et al.* Inhibition of proteolysis and cell-cycle progression in a multiubiquitination-deficient yeast mutant. *Mol. Cell. Biol.* **14**, 5501–5509 (1994).
- Chi, Y. *et al.* Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev.* **15**, 1078–1092 (2001).
- Adams, J. & Kauffman, M. Development of the proteasome inhibitor Velcade (bortezomib). *Cancer Invest.* **22**, 304–311 (2004).
- Mitsiades, N. *et al.* Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proc. Natl Acad. Sci. USA* **99**, 14374–14379 (2002).
- Fleming, J. A. *et al.* Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. *Proc. Natl Acad. Sci. USA* **99**, 1461–1466 (2002).
- Hinnebusch, A. G. & Natarajan, K. Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. *Eukaryot. Cell* **1**, 22–32 (2002).
- Kornitzer, D., Raboy, B., Kulkarni, R. G. & Fink, G. R. Regulated degradation of the transcription factor GCN4. *EMBO J.* **13**, 6021–6030 (1994).
- Albrecht, G., Mosch, H. U., Hoffmann, B., Reusser, U. & Braus, G. H. Monitoring the Gcn4 protein-mediated response in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 12696–12702 (1998).
- Hilt, W., Enenkel, C., Gruhler, A., Singer, T. & Wolf, D. H. The PRE4 gene codes for a subunit of the yeast proteasome necessary for peptidylglutamyl-peptide-hydrolyzing activity. Mutations link the proteasome to stress- and ubiquitin-dependent proteolysis. *J. Biol. Chem.* **268**, 3479–3486 (1993).
- Salghetti, S. E., Caudy, A. A., Chenoweth, J. G. & Tansey, W. P. Regulation of transcriptional activation domain function by ubiquitin. *Science* **293**, 1651–1653 (2001).
- Kim, S., Herbst, A., Tworkowski, K., Salghetti, S. & Tansey, W. Skp2 regulates Myc protein stability and activity. *Mol. Cell* **11**, 1177–1188 (2003).
- von der Lehr, N. *et al.* The F-Box protein Skp2 participates in c-Myc proteasomal degradation and acts as a cofactor for c-Myc-regulated transcription. *Mol. Cell* **11**, 1177–1188 (2003).
- Aparicio, O., Geisberg, J. & Struhl, K. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) 21.3.1–21.3.12 (Wiley, New York, 2004).
- Stitzel, M. L., Durso, R. & Reese, J. C. The proteasome regulates the UV-induced activation of the AP-1-like transcription factor Gcn4. *Genes Dev.* **15**, 128–133 (2001).
- Reid, G. *et al.* Cyclic, proteasome-mediated turnover of unliganded and liganded ER α on responsive promoters is an integral feature of estrogen signalling. *Mol. Cell* **11**, 695–707 (2003).
- Gonzalez, F., Delahodde, A., Kodadek, T. & Johnston, S. A. Recruitment of a 19S proteasome subcomplex to an activated promoter. *Science* **296**, 548–550 (2002).
- Morris, M. C. *et al.* Cks1-dependent proteasome recruitment and activation of CDC20 transcription in budding yeast. *Nature* **423**, 1009–1013 (2003).
- Verma, R. *et al.* Deubiquitination and degradation of proteins by the 26S proteasome requires the Rpn11 metalloprotease motif. *Science* **298**, 611–615 (2002).
- Muratani, M., Kung, C., Shokat, K. M. & Tansey, W. P. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**, 887–899 (2005).
- Hirst, M., Kobor, M. S., Kuriakose, N., Greenblatt, J. & Sadowski, I. GAL4 is regulated by the RNA polymerase II holoenzyme-associated cyclin-dependent protein kinase SRB10/CDK8. *Mol. Cell* **3**, 673–678 (1999).
- Guthrie, C. & Fink, G. R. *Guide to Yeast Genetics and Molecular Biology* (Academic, San Diego, 1991).
- Drysdale, C. M. *et al.* The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol. Cell. Biol.* **15**, 1220–1233 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Finley, D. H. Wolf, R. Young, A. Hinnebusch, J. Shaw, B. Westermann, J. Nunnari and G. Braus for gifts of strains and reagents; B. Tansey for communicating results before publication; and J. Shaw, B. Westermann, J. Nunnari, S. Sadis, A. Ansari and the members of the Deshaies laboratory for comments and criticism. This work was supported in part by an NIH Research Project Grant to R.J.D. R.J.D. is an Investigator of the Howard Hughes Medical Institute. J.R.L. was supported by an NIH National Research Service Award and a Caltech–Amgen Postdoctoral Fellowship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.J.D. (deshaies@caltech.edu).

Intrinsic dynamics of an enzyme underlies catalysis

Elan Z. Eisenmesser¹, Oscar Millet^{2,†}, Wladimir Labeikovsky¹, Dmitry M. Korzhnev², Magnus Wolf-Watz^{1,†}, Daryl A. Bosco^{1,†}, Jack J. Skalicky^{3,†}, Lewis E. Kay² & Dorothee Kern¹

A unique feature of chemical catalysis mediated by enzymes is that the catalytically reactive atoms are embedded within a folded protein. Although current understanding of enzyme function has been focused on the chemical reactions and static three-dimensional structures, the dynamic nature of proteins has been proposed to have a function in catalysis^{1–5}. The concept of conformational substates has been described⁶; however, the challenge is to unravel the intimate linkage between protein flexibility and enzymatic function. Here we show that the intrinsic plasticity of the protein is a key characteristic of catalysis. The dynamics of the prolyl *cis*–*trans* isomerase cyclophilin A (CypA) in its substrate-free state and during catalysis were characterized with NMR relaxation experiments. The characteristic enzyme motions detected during catalysis are already present in the free enzyme with frequencies corresponding to the catalytic turnover rates. This correlation suggests that the protein motions necessary for catalysis are an intrinsic property of the enzyme and may even limit the overall turnover rate. Motion is localized not only to the active site but also to a wider dynamic network. Whereas coupled networks in proteins have been proposed previously^{3,7–10}, we experimentally measured the collective nature of motions with the use of mutant forms of CypA. We propose that the pre-existence of collective dynamics in enzymes before catalysis is a common feature of biocatalysts and that proteins have evolved under synergistic pressure between structure and dynamics.

The most basic principle of enzyme catalysis is the ability of an enzyme to decrease the transition-state energy, thereby catalysing the chemical reaction. A wealth of information about the kinetics and thermodynamics of enzyme-catalysed reactions has been obtained by monitoring the conversion of substrates into products. However, much less is known about the kinetics and energetics of conformational processes in the protein. Because it is the protein component that alters the transition state energy, enzyme function depends on transitions from ground states to higher-energy states of the enzyme and the reactant. A detailed understanding of the entire trajectory of catalysis is therefore a current challenge. Structures of intermediates provide snapshots of conformational changes, and a set of experimental methods, such as fluorescence, mass spectroscopy and NMR, have been developed to monitor the kinetics of these changes. We have recently provided a proof of principle that enzyme dynamics can be monitored during catalysis at multiple sites by NMR relaxation experiments with the use of cyclophilin A (CypA)⁵. This enzyme belongs to the family of prolyl-isomerases that catalyse the *cis*–*trans* isomerization of prolyl peptide bonds. Although CypA is involved in a series of biomedically important processes¹¹, its natural biological role and mechanism of action are still in debate. CypA is the target of the

immunosuppressive drug cyclosporin A (CsA) and is essential for HIV-1 virulence¹².

Here we use new NMR relaxation dispersion experiments¹³ to compare motions in free CypA with those during turnover. We show that protein dynamics associated with catalysis is a built-in property of the enzyme that is also manifested in the free protein. The motions are collective, propagating from the active site to remote sites. The results show that intrinsic plasticity on a basic structural template is a crucial element for catalytic function, and that proteins have therefore evolved under synergistic pressure between structure and dynamics.

Relaxation dispersion experiments probe molecular motions in the microsecond to millisecond timescale quantitatively and with much higher sensitivity than traditional transverse relaxation experiments¹³. The additional line-broadening of NMR signals caused by conformational exchange between two states (R_{ex}) depends on the sum of forward and reverse rates of interconversion (k_{ex}), the relative populations of the exchanging species (p_A and p_B) and the chemical shifts between the exchanging species ($\Delta\omega$), with

$$R_{ex} = p_A p_B \frac{\Delta\omega^2}{k_{ex}} \left(1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh \frac{k_{ex}}{4\nu_{CPMG}} \right) \quad (1)$$

in the fast exchange limit¹⁴.

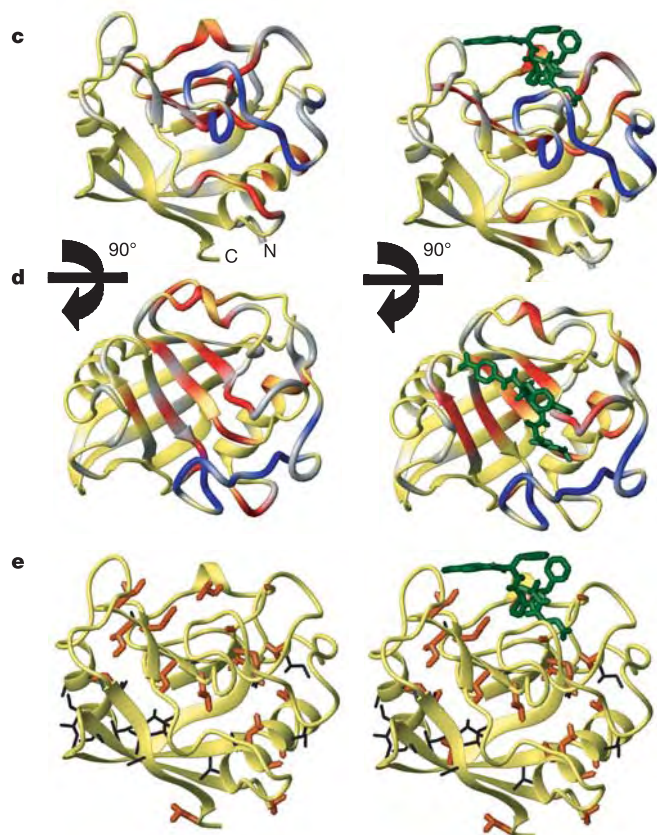
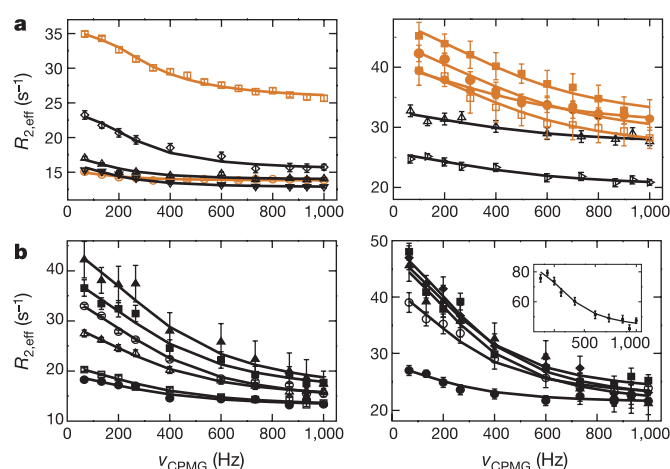
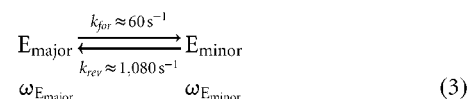
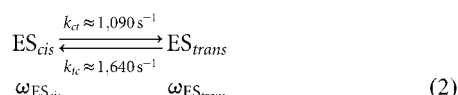
The dependence of R_{ex} on the applied external field (ν_{CPMG} , in which CPMG stands for Carr–Purcell–Meiboom–Gill) is the key element in relaxation dispersion experiments (Fig. 1a, b). It is noted that an atom with R_{ex} contributions reports on a change in its electronic environment, but this does not necessarily represent physical movement of this atom.

We applied this method to ¹⁵N-labelled CypA catalysing the *cis*–*trans* isomerization of *N*-succinyl-Ala-Phe-Pro-Phe-*p*-nitroanilide (Fig. 1, right). The high sensitivity of these experiments leads to the identification of many more amides in CypA with conformational exchange than originally described⁵. Clear and quantitative analysis of the relaxation data require conditions of two-state exchange. However, CypA interconverts between at least three states during the catalytic cycle: the free enzyme (E) and the two enzyme–substrate (ES) complexes with substrate bound in the *cis* and *trans* conformations. The system was therefore biochemically ‘tuned’ to two-state exchange: by using an excess of substrate at 10 °C, 95% saturation with substrate could be reached in which the dispersion profiles detect only two-state conformational exchange corresponding to the catalytic step of *cis*–*trans* isomerization (scheme (2)). Under these conditions, contributions to R_{ex} from substrate binding and dissociation are negligible (see Supplementary Information).

Individual fits of dispersion profiles on a collection of 30 amides resulted in similar rate constants, a remarkable result that justifies a

¹Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA. ²Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ³National High Magnetic Field Laboratory at Florida State University, Tallahassee, Florida 32310, USA. [†]Present addresses: Plataforma de Biomoléculas, Parc Científic de Barcelona, Josep Samitier 1–5, 08028 Barcelona, Catalonia, Spain (O.M.); The Scripps Research Institute, Department of Chemistry, La Jolla, California 92037, USA (D.A.B.); University of Utah School of Medicine, Department of Biochemistry, Salt Lake City, Utah 84132, USA (J.J.S.); University of Umeå, Department of Biochemistry, Umeå SE-901 87, Sweden (M.W.W.).

global fit with a single rate constant, k_{ex} (Fig. 1a, b, right). This result is further buttressed by ^{13}C -CPMG dispersion experiments on methyl side chains¹³ (Fig. 1a, right). Comparison of the dynamic hotspots of amide nitrogens (Fig. 1c, d) and methyl carbons (Fig. 1e) reveals an extended dynamic network emanating from the active site of CypA. ^{15}N and ^{13}C relaxation data could be globally fitted with the same k_{ex} value of $2,730 \pm 220 \text{ s}^{-1}$ (mean $\pm$ s.d.) and separation into the individual rate constants (scheme (2)) was based on the relative populations of *cis* and *trans* bound substrates determined previously¹⁵. This k_{ex} value of protein motion is very close to the sum of the rate constants of *cis*-to-*trans* (k_{ct}) and *trans*-to-*cis* (k_{tc}) substrate isomerization on the enzyme ($k_{\text{ex}} = 2,500 \pm 500 \text{ s}^{-1}$; mean $\pm$ s.d.)¹⁵.



Our earlier work indicated a possible intimate correlation between motion on the enzyme and substrate turnover⁵, but the system was underdetermined with respect to the exact rate of conformational exchange. The present results show explicitly that conformational changes of the enzyme coincide with substrate turnover.

Although it is not surprising that conformational rearrangements occur in the working enzyme, we found that most of these same residues exhibit exchange in the resting, substrate-free state (Fig. 1, left). Quantitative analysis revealed that most residues can be globally fitted with a k_{ex} value of $1,140 \pm 200 \text{ s}^{-1}$ (mean $\pm$ s.d.) (Fig. 1a, left) for an exchange process between a major and a minor population (of less than 10%; see Methods). Notably, one of the exchange rates in free CypA is comparable to those observed during turnover:

It is also noteworthy that the calculated chemical-shift differences (equation (1)) between E_{major} and E_{minor} ($\omega_{\text{E}_{\text{major}}} - \omega_{\text{E}_{\text{minor}}}$; scheme (3)) are similar to those between ES_{cis} and ES_{trans} ($\omega_{\text{ES}_{\text{cis}}} - \omega_{\text{ES}_{\text{trans}}}$; scheme (2)) for residues remote from the substrate, suggesting that the *cis* substrate binds E_{major} to form ES_{cis} , whereas the *trans* substrate binds E_{minor} to form ES_{trans} . This would imply that binding of the substrates shifts the pre-existing, highly skewed equilibrium to a more balanced one.

Besides the interconversion detected between E_{major} and E_{minor} , a specific region (the loop comprising residues 65–84) clearly moves faster, with a k_{ex} of $2,260 \pm 200 \text{ s}^{-1}$ (mean $\pm$ s.d.) (Fig. 1b, left) in the free enzyme. This result implies that motions in proteins are usually more complex than mere interconversion between two conformational states. Folded proteins must in fact be described with a multidimensional energy landscape. However, with substrate bound, this loop seems to fluctuate collectively with the rest of the protein as catalysis occurs.

The fact that in free CypA all residues within the two groups can be

Figure 1 | Protein dynamics necessary for catalysis is an intrinsic property of the enzyme. Quantitative analysis of protein dynamics of resting (left) and working (right) CypA. **a, b**, Global fits of NMR relaxation dispersion data³⁰ using both ^{15}N backbone amide (black) and ^{13}C methyl groups (orange) of CypA in the absence of substrate (left) and during turnover (right) are shown. Residues in free CypA had to be globally fitted in two groups with $k_{\text{ex}} = 1,140 \pm 200 \text{ s}^{-1}$ for group I (**a**) and $k_{\text{ex}} = 2,260 \pm 200 \text{ s}^{-1}$ (means $\pm$ s.d.) for group II (**b**), whereas during turnover the distinction of rates was not apparent and consequently all residues were globally fitted together with $k_{\text{ex}} = 2,730 \pm 220 \text{ s}^{-1}$ (mean $\pm$ s.d.) (**a, b**, right panel). Residues Leu³⁹C δ 1 (open square), Arg⁵⁵N^H (open triangles, point sideways), Ile⁵⁷C γ 2 (filled squares), Leu⁹⁸C δ 1 (open circles), Ser⁹⁹N^H (open uptriangles), Ala¹⁰³C β (filled circles), Gly¹¹⁰N^H (open diamonds) and Glu¹²⁰N^H (open downtriangles); and residues Asp⁶⁶N^H (filled circles), Phe⁶⁷N^H (filled squares), Thr⁶⁸N^H (filled diamonds), Asn⁷¹N^H (filled uptriangle), Gly⁷⁴N^H (open circles), Lys⁷⁶N^H (open square), Ser⁷⁷N^H (open uptriangle) and Lys⁸²N^H (inset) are shown for groups I and II, respectively. **c, d**, Amides undergoing chemical exchange in free CypA and during turnover of the substrate *N*-succinyl-Ala-Phe-Pro-Phe-*p*-nitroanilide (green) are coloured in red and blue according to group I and II, respectively. All residues of group II are located in a loop region (blue). Dynamics of amides shown in grey could not be characterized because of the absence of the amide (Pro) or peak overlap. Dynamics data are plotted on the crystal structures of free CypA and CypA bound to the *cis* conformer of a similar substrate¹⁶, shown in two views (**c** and **d**) differing by a 90° horizontal rotation. **e**, Methyl side chains exhibiting chemical exchange during catalysis are shown in orange; those that do not show dispersion are shown in black. For each relaxation dispersion curve, two duplicate points were collected and used to estimate the absolute uncertainties together with the signal-to-noise ratio of each individual spectrum.

fitted with single rate constants is indicative of networks with collective motions, but it is not proof. As a direct test of the collective nature of the dynamics and to identify residues that build a common network, the system was locally disturbed by mutations within the network, and the effects of these mutations on the structure and dynamics were examined. We compared mutations in the active site—R55A, R55K and F113W—with amino-acid substitutions within loops adjacent to the active site, H70A and K82A (Fig. 2). These residues were chosen because they are located at different positions within the dynamic network. In addition, Lys 82 is drastically affected by substrate binding⁵ despite the fact that it is more than 10 Å from the substrate in the crystal structure¹⁶. Catalytic activities, given as k_{cat}/K_m (ref. 17), were severely diminished for the active-site mutants (R55A, $0.09 \pm 0.01\%$; R55K, $0.1 \pm 0.01\%$; F113W, $0.03 \pm 0.01\%$), whereas mutations in the loop did not markedly alter enzymatic activities (H70A, $108 \pm 10\%$; K82A, $88 \pm 7\%$). All mutants except H70A had lower substrate affinities, as determined from chemical-shift changes as a function of substrate concentration. Mutation of Arg 55 results in greatly decreased activity because this residue lowers the activation barrier of *cis-trans* isomerization by decreasing the double-bond character of the prolyl peptide bond through hydrogen bonding to the proline nitrogen.

All mutations cause chemical-shift changes, indicating changes in chemical environment that propagate throughout much of the protein. Moreover, it does not matter from which ‘end’ the system is disturbed; changes are always seen in a common network. By cross-correlating the chemical-shift changes caused by the mutations (Fig. 2d–f), we identified networks between sites. Remarkably, these networks are similar to those identified by conformational exchange on the wild-type enzyme (Fig. 1c–e).

What is the physical reason for this similarity? ^{15}N relaxation dispersion experiments on all mutants showed not only chemical exchange at a similar set of residues to those in the wild type but also similar exchange rates (k_{ex} ; Supplementary Table S1). However, the amplitudes of the dispersion profiles were changed by constant factors for nearly all amides for each mutant (Fig. 3a). This must reflect a shift in relative populations of the two exchanging states because k_{ex} remains the same (equation (1)). It also implies that the chemical shifts between the two states ($\Delta\omega$), and hence their structural differences, are identical for wild-type CypA and all CypA mutants (scheme (3)); Supplementary Table S2). Consequently, the dynamics data of the mutants provide direct evidence for the

collectiveness of motions and show that the mutations shift a highly skewed conformational equilibrium that is already present in wild-type CypA. Although most mutations such as H70A and K82A shift the equilibrium towards the minor state, R55A and binding of the inhibitor CsA have the opposite effect. In fact, the equilibrium is so far shifted towards the major conformation in the presence of CsA (more than 98%) that dispersion could be detected only for those amides with exceptionally large chemical-shift differences. A suppression of chemical exchange in CypA by CsA has previously been described qualitatively¹⁸. We now state the physical underpinning for this phenomenon: that inhibitor binding shifts a pre-existing equilibrium towards the major conformation. The observed global response to mutagenic perturbations further suggests that the dynamics of the loop region, although exhibiting a different exchange rate, is correlated with the dynamics of the core. Structurally this makes sense because Phe 67 within the loop region is part of a hydrophobic cluster that includes Leu 39, Phe 46, Phe 48 and Ile 78.

To corroborate the collective nature of the conformational transition, we performed an additional type of measurement on all amides. In this experiment, the sign of the changes in ^{15}N chemical shift between the major and minor states were determined for wild-type and mutant forms of CypA by measuring ^{15}N chemical-shift differences, Ω_N , between heteronuclear single-quantum coherence (HSQC) and heteronuclear multiple-quantum coherence (HMQC) spectra (Fig. 3b). Ω_N depends on the position in the ^{15}N dimension of the minor form with respect to the major form and on the relative populations¹⁹. For all amides with measurable differences, the signs are the same in all CypA forms. The amount of shift for all detected amides is proportional to the population differences determined from the relaxation dispersion experiments (Fig. 3c). This independent experiment further supports our original observation of collective dynamics in the free enzyme. Furthermore, ^{15}N chemical shifts of the minor conformation are not that of an unfolded protein²⁰, revealing the sampling of folded substates.

We have thus shown that the conformational exchange in CypA detected during catalysis already exists in the absence of substrate. In other words, the enzyme’s intrinsic dynamic ‘personality’ is well suited to its catalytic role. Moreover, the frequencies of these intrinsic motions are almost identical in the resting and turning-over enzyme and coincide with the overall turnover number. This correspondence might be of wide-ranging importance, given that in many enzymes, such as triosephosphate isomerase^{21,22}, dihydrofolate reductase^{23,24},

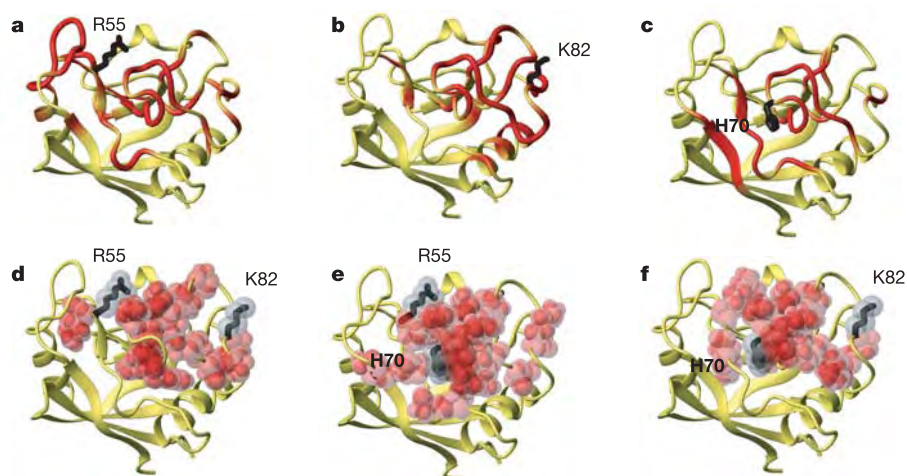


Figure 2 | Identification of residues that build a common dynamic network in CypA. **a–c**, Amide resonances for R55A (**a**), K82A (**b**) and H70A (**c**) that show changes in chemical shift relative to wild-type CypA are coloured in

red. **d–f**, Common residues with chemical-shift changes in R55A and K82A (**d**), R55A and H70A (**e**) and H70A and K82A (**f**) are displayed in red as van der Waals radii, delineating the collective network.

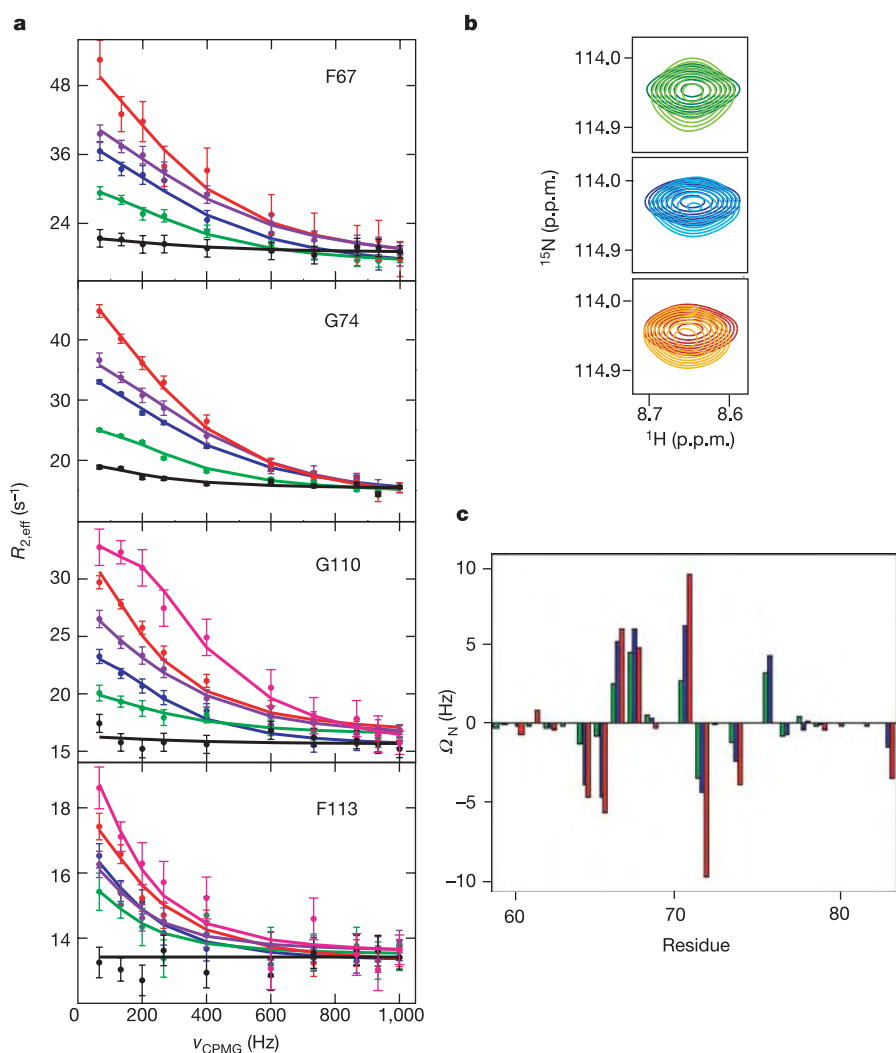


Figure 3 | Probing correlated motions of CypA by point mutations. **a**, CPMG relaxation dispersion data of four representative ^{15}N backbone amides are shown for the wild type (blue), R55K mutant (violet), R55A mutant (green), K82A mutant (red), H70A mutant (magenta) and wild-type CypA bound to CsA (black). Although the k_{ex} values for all forms are similar, the amplitudes of the dispersion profiles, which are proportional to R_{ex} , differ between the forms by a constant factor directly reflecting changes in populations of the two exchanging states. **b**, **c**, Mutations shift the populations of exchanging states as established by a comparison of chemical shifts of the major conformer in HSQC (darker colour) and HMQC (lighter colour) spectra¹⁹, visualized for Gly 74 in R55A (green), wild type (blue) and K82A (orange). **b**, The less populated state is shifted upfield in the ^{15}N dimension relative to the major conformation for all proteins tested. **c**, The differences in ^{15}N chemical shift between HSQC and HMQC spectra (Ω_{N}) (ref. 19) for residues 60–83 are shown for R55A (green), wild type (blue) and K82A (red). The relative changes in Ω_{N} between the different forms of CypA are proportional to the changes in population¹⁹ determined from the dispersion profiles in **a**. Error analysis was performed as described in the legend to Fig. 1.

HIV-1 protease²⁵, RNase A²⁶ and many others¹⁴, motions with rates of the same order of magnitude as the turnover number have been detected in their free or ligand-bound state. The similarity of dynamics in free and catalytically active CypA, together with these data on other enzymes, indicates that in enzymes with a fast turnover, catalytic power might be limited by rates of conformational rearrangements.

Our results provide further experimental evidence that the observed fluctuations at equilibrium are a result of a concerted global process. The identification of an extended dynamic network in CypA illustrates how proteins can propagate local changes over large distances. Coupled networks in proteins have previously been proposed on the basis of computational techniques^{3,7,8} and sequence-based methodologies⁹, and recent computational studies on CypA during substrate turnover have suggested correlated dynamic networks that largely agree with our experimental findings¹⁰.

Although the existence of motions in proteins is a fundamental thermodynamic concept, our results show pre-sampling of conformational substates before catalysis that are harvested for catalytic turnover. This is directly analogous to recent findings of the role of pre-existing equilibria for allosteric signalling and ligand binding^{14,27}. A picture is therefore emerging in which conformational events that occur during protein function are already present before ligands bind. In contrast to being folded in a single native conformation, proteins have evolved to sample multiple defined conformations that are critical for function. The concept of conformational substates and exploration of the energy landscape of proteins was put forward 30

years ago by Frauenfelder and collaborators⁶; the challenge now is to understand the connections between structure, energy landscapes, dynamics and function.

METHODS

Materials and enzyme activity. Wild-type and mutant proteins were expressed and purified as described previously⁵. For CypA containing [^{13}C]methyl-labelled CypA was produced as described previously²⁸.

Unless otherwise noted, NMR samples contained 0.7 mM CypA in 50 mM Na_2HPO_4 pH 6.5, and 1 mM dithiothreitol with 10% $^2\text{H}_2\text{O}$. The peptide *N*-succinyl-Ala-Phe-Pro-Phe-*p*-nitroanilide was used for NMR studies of CypA dynamics during catalysis and for activity assays by the coupled chymotrypsin assay at 10 °C (ref. 17). At least three measurements were used to determine $k_{\text{cat}}/K_{\text{m}}$ values and their respective uncertainties. For saturated NMR samples 3.2 mM peptide and 3.5 mM CsA were used.

NMR spectroscopy and data analysis. All spectra were collected on Varian 600-MHz and 800-MHz spectrometers. Chemical shifts for CypA alone and in the presence of CsA were published previously¹⁸. Chemical shifts for CypA in the presence of substrate were determined with substrate titration data and confirmed by three-dimensional nuclear Overhauser enhancement spectroscopy-HSQC, a CBCA(CO)NH and a HNCACB spectrum (where CA is the alpha carbon and CB is the beta carbon).

Unless noted otherwise, all ^{15}N CPMG relaxation data were collected at 10 °C and 600 MHz. This lower temperature was used because of the much higher binding affinity for the substrate than at 25 °C (refs 5, 15). Data were collected on free CypA at 25 °C with protein concentrations of 1 and 2 mM to rule out sample aggregation effects on chemical exchange (Supplementary Fig. S1). Amide spectra were collected with TROSY (transverse relaxation-optimized spectroscopy)-CPMG sequences^{13,29}, applying 10–14 CPMG field strengths

ranging from 50 to 1,000 Hz. Methyl side-chain CPMG pulse sequences¹³ were performed at 10 °C at 600 and 800 MHz.

Relaxation dispersion data were based on peak intensities and analysed with the generalized Carver–Richard equation for a two-site exchange³⁰ with the use of in-house scripts. For each relaxation dispersion curve two duplicate points were collected and used to estimate the absolute uncertainties together with the signal-to-noise ratio of each individual spectrum. Dispersion curves were then fitted individually to a two-site exchange for all residues exhibiting chemical exchange.

For CypA in the presence of saturating substrate concentration, the individual fits gave exchange rates between 2,400 and 3,000 s⁻¹ justifying a global fit. For free CypA, individual fits clustered in two groups with rates in the ranges 900–1,400 s⁻¹ for group I and 2,000–3,000 s⁻¹ for group II. The separation of residues into these two groups was strongly buttressed by dispersion experiments at 25 °C, at which temperature the difference in rates between the two groups is more pronounced (Supplementary Fig. S2). Residues within those groups were fitted globally. Uncertainties for all global fits were calculated with a jackknife approach (Supplementary Table S1).

The similarity of exchange rates extracted from fits of dispersion profiles on a per-residue basis supports a two-state model of exchange. To establish further that this is true we compared a series of point mutants. Changes in global amplitude by constant factors in relaxation dispersion profiles for the mutant forms of CypA (Fig. 3a and Supplementary Table S3) with nearly identical exchange rates can be interpreted only by a shift in population between two states. Exact absolute populations cannot be determined because the process is in the fast exchange limit for most resonances and the chemical-shift difference is not known. However, data for the wild type and mutants can be fitted only with highly skewed populations (Supplementary Table S2). The observed k_{ex} and R_{ex} values define an upper limit of about 7% for the minor conformation of wild-type CypA because larger populations would yield $\Delta\omega$ values (equation (1)) that were so small as to result in unobservable chemical-shift changes in ¹⁵N HSQC experiments for the mutants. The result of highly skewed populations is further buttressed by simultaneous fits of [¹³C]methyl dispersion data at 600 and 800 MHz of residues with exchange in the intermediate time regime ($\alpha < 1.5$; ref. 14), for which p_{minor} was calculated to be 5%.

HMQC and HSQC spectra were collected at 10 °C and 600 MHz for wild-type and mutant forms of CypA with the use of recently published pulse sequences¹⁹. Chemical shifts of the minor enzyme form ($\omega_{E_{minor}}$) were calculated by using $\Delta\omega$ obtained from the dispersion experiments (equation (1)) together with the sign of $\Delta\omega$ determined from the HMQC/HSQC experiment. Although the relation between the frequency differences of corresponding correlations in HMQC and HSQC data sets in the general case of an exchanging system is complex (see for example, equation (5) of ref. 19), simulations establish that the correlation between the difference and the population of the minor state, p_B , is nearly linear for $p_B \leq 0.15$. Thus, in an exchange involving highly skewed populations, the relative shifts of HMQC and HSQC correlations as a function of a perturbation that does not change the rate and shift differences between states can be used as a direct measure of changes in p_B .

Received 27 June; accepted 3 August 2005.

- Jencks, W. P. *Catalysis in Chemistry and Enzymology* (Dover, New York, 1987).
- Hammes, G. Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221–8228 (2002).
- Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196–1202 (2003).
- García-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How enzymes work: analysis by modern rate theory and computer simulations. *Science* **303**, 186–195 (2004).
- Eisenmesser, E. Z., Akke, M., Bosco, D. A. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520–1523 (2002).
- Austin, R. H. *et al.* Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
- Berendsen, H. J. C. & Hayward, S. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **10**, 165–169 (2000).
- Rod, T. H., Radkiewicz, J. L. & Brooks, C. L. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl Acad. Sci. USA* **100**, 6980–6985 (2003).
- Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69 (2003).
- Agarwal, P. K., Geist, A. & Gorin, A. Protein dynamics and enzymatic catalysis: Investigating the peptidyl-prolyl *cis-trans* isomerization activity of cyclophilin A. *Biochemistry* **43**, 10605–10618 (2004).
- Schmid, F. X. Prolyl isomerases. *Adv. Protein Chem.* **59**, 243–282 (2001).
- Goff, S. P. Genetic control of retrovirus susceptibility in mammalian cells. *Annu. Rev. Genet.* **38**, 61–85 (2004).
- Mulder, F. A. A., Mittermaier, A., Hon, B., Kahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nature Struct. Biol.* **8**, 932–935 (2001).
- Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).
- Kern, D., Kern, G., Scherer, G., Fischer, G. & Drakenberg, T. Kinetic analysis of cyclophilin-catalyzed prolyl *cis/trans* isomerization by dynamic NMR spectroscopy. *Biochemistry* **34**, 13594–13602 (1995).
- Zhao, Y. & Ke, H. Crystal structure implies that cyclophilin predominantly catalyzes the *trans* to *cis* isomerization. *Biochemistry* **35**, 7356–7361 (1996).
- Fischer, G., Bang, H. & Mech, C. Determination of enzymatic catalysis for the *cis-trans*-isomerization of peptide binding in proline-containing peptides. *Biomed. Biochim. Acta* **43**, 1101–1111 (1984).
- Ottiger, M., Zerbe, O., Guntert, P. & Wuthrich, K. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* **272**, 64–81 (1997).
- Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. Reconstructing NMR spectra of ‘invisible’ excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* **124**, 12352–12360 (2002).
- Wishart, D. S., Sykes, B. D. & Richards, F. M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333 (1991).
- Rozovsky, S., Jogl, G., Tong, L. & McKernott, A. E. Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *J. Mol. Biol.* **310**, 271–280 (2001).
- Williams, J. C. & McDermott, A. E. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry* **34**, 8309–8319 (1995).
- Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 119–140 (2004).
- McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J. & Wright, P. E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl Acad. Sci. USA* **102**, 5032–5037 (2005).
- Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M. & Torchia, D. A. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Struct. Fold. Des.* **7**, 1047–1055 (1999).
- Cole, R. & Loria, J. P. Evidence for flexibility in the function of ribonuclease A. *Biochemistry* **41**, 6072–6081 (2002).
- Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behaviour in a single-domain signalling protein. *Science* **291**, 2429–2433 (2001).
- Rosen, M. K. *et al.* Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263**, 627–636 (1996).
- Loria, J. P., Rance, M. & Palmer, A. G. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J. Biomol. NMR* **15**, 151–155 (1999).
- Carver, J. P. & Richards, R. E. A general two-site solution for the chemical exchange produced dependence of T₂ upon the Carr–Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105 (1972).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. H. Wildman for discussions. This work was supported by NIH grants to D.K., by a grant from the Canadian Institutes of Health Research to L.E.K., and by a grant from the Swedish Research Council to M.W.V. Part of the NMR studies was performed at the NHMFL at Florida with support from the NSF.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.K. (dkern@brandeis.edu).

CORRIGENDUM

doi:10.1038/nature04289

Eocene bipolar glaciation associated with global carbon cycle changes

Aradhna Tripathi, Jan Backman, Henry Elderfield & Patrizia Ferretti

Nature 436, 341–346 (2005)

We wish to clarify that mass accumulation rates in this Article were calculated using dry bulk densities from Michael Vanden Berg (personal communication) and that the method used for calculating the calcite compensation depth (CCD) in Fig. 1a (linear extrapolation of the CCD) is detailed in a forthcoming publication¹.

Also, the dark green and dark grey lines in Fig. 2 of the Article should have contained symbols to distinguish between data from different laboratories for site 1218, and a revised version of Fig. 2 is accordingly shown here. Benthic foraminiferal data from our study are now indicated by open circles (dark colours, site 1218; light colours, site 1209) and published data² for site 1218 are represented as crosses.

Our conclusions remain unchanged.

1. Lyle, M., Olivarez-Lyle, A., Backman, J. & Tripathi, A. Biogenic sedimentation in the Eocene Equatorial Pacific—the stuttering greenhouse and Eocene carbonate compensation depth. *Proc. ODP Sci. Res.* (in the press).
2. Lear, C. H., Rosenthal, Y., Coxall, H. K. & Wilson, P. A. Late Eocene to early Miocene ice sheet dynamics and the global carbon cycle. *Paleoceanography* 19, doi:10.1029/2004PA1039 (2004).

ERRATUM

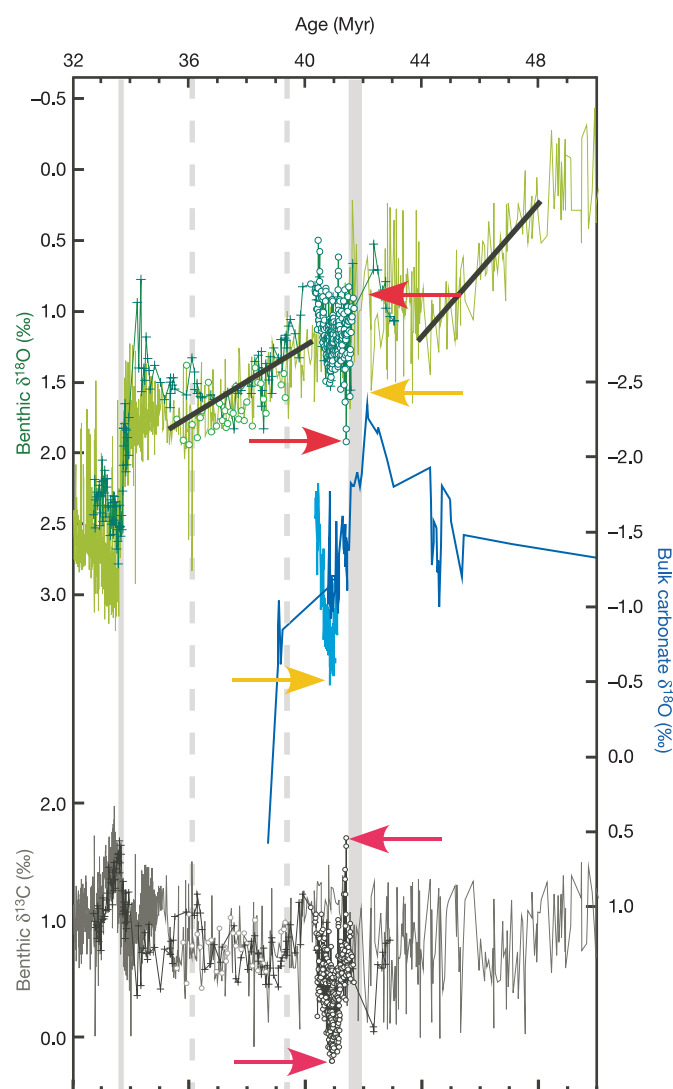
doi:10.1038/nature04265

Marine microorganisms and global nutrient cycles

Kevin R. Arrigo

Nature 437, 349–455 (2005)

In this Review, the digital object identifier (DOI) number was incorrectly given as doi:10.1038/nature04158. The correct DOI number for this Review is doi:10.1038/nature04159.



-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

A transparent process

The city of Trieste is bucking the trend for scientific expansion in Italy. Its success is encapsulated neatly in a telling statistic. In Trieste there are 37 scientists per 1,000 members of its population compared with a level of just five for Italy as a whole, and six in the European Union (see page 124).

What accounts for such a disparity? One of the major factors is that few of the science jobs in Trieste arise from Italy's state-funded university system. The state system has two significant problems. First, funding is linked to the state budget, which for science has been relatively flat for several years. Second, unlike Trieste's international research institutes, state universities run slow *concorsi* recruitment competitions to fill their positions. The *concorsi* system has the added disadvantage of lacking transparency and, as a result, invites cronyism.

Trieste's success highlights a problem that stretches beyond Italy into the rest of Europe, where many state-run university systems have similar issues. In Germany, under the *Habilitation* system, which is slowly being phased out, young researchers are disproportionately

beholden to their mentors. In France, major research agencies often appoint young scientists to permanent positions fresh from PhD programmes, leaving others who are equally skilled bewildered — and pursuing jobs outside the country. And in Spain, new regional programmes are trying to establish systems that give PhDs from outside Spain an opportunity for employment (see *Nature* **428**, 448–449; 2004).

Recruiters in Europe would be wise to emulate transparent systems that give candidates from all over the world a fair shot. And job-seekers would do well to seek out opportunities in such systems and be wary of universities that lack a transparent recruiting practice. Maybe then Trieste's high proportion of scientists would become less of an anomaly and more of a norm.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 345 Park Avenue South,
 10th Floor, New York,

NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Hendricus Hoogenboom, chief scientific officer, Ablynx, Ghent, Belgium



2002-05: Chief scientific officer, Merus, Utrecht, the Netherlands

1999-2002: Senior vice-president of discovery research, Dyax, Liege, Belgium

1997-99: Founder and director, Target Quest, Maastricht

1995-2001: Associate professor of medical biotechnology, University of Maastricht, the Netherlands

Hendricus Hoogenboom is thinking small. After 20 years of juggling academic and commercial work in antibody engineering, his latest job takes him into the emerging field of 'nanobodies', antibody fragments that are smaller, more stable and easier to tailor to specific targets.

Among his 20-plus patents are those on Humira, the first fully human antibody approved for therapy. Seeing that many problems with conventional antibodies remain unsolved after years of engineering, Hoogenboom feels that Ablynx's technology offers a possible solution.

But even if a technology is promising, make sure there's a market for it, he warns would-be entrepreneurs.

"In Europe, many academics are seduced by grant money to initiate commercial activities," he says. "My advice is to ignore government grants until your idea has been truly tested by venture capitalists and local biotech experts."

He speaks from experience. He co-founded the Dutch biotech Merus with a plan to produce cocktails of human antibodies, combining the simplicity of monoclonal antibodies (derived from just one cell line) with the potency of polyclonals. It turned out to be difficult to raise enough funds for this venture.

"That's probably because the industry is hesitant to get into more complex biopharmaceuticals, even if the potency can be improved drastically," Hoogenboom says. "Be prepared to accept that you can be too far ahead of the industry." Ablynx, however, has plenty of funding and is already planning clinical trials of its first lead molecule.

After gaining his PhD in agricultural sciences at the Catholic University of Leuven in Belgium, Hoogenboom crossed the channel for a postdoc at the UK Medical Research Council in Cambridge. Working with "a great mentor", Gregory Winter, he co-invented phage display technology for the isolation and engineering of human antibodies. That led to two years at Winter's biotech Cambridge Antibody Technology.

The urge to work on cancer took Hoogenboom back to academia. But realizing how long it would be before this research would benefit patients pushed him to set up his first company, Target Quest. Having explored R&D in both camps, he has now decided to stick with industry.

"Drug and technology development in industry are team sports, which is often very different from the working style in academic labs," he says. "If you can adjust to this after a PhD, you are likely to do well in the biotech industry."

Janet Wright

SCIENTISTS & SOCIETIES

Retreat to make progress

Most postdocs in the Netherlands want to stay in academia, but only a few will be able to do so. In 1999, Peter Peters, dean of postdoc affairs at the Netherlands Cancer Institute, helped postdocs to establish an annual retreat focusing on career development. Since then, attendance has doubled to 150. The theme of last month's retreat was 'Making the right moves', based on a Burroughs Wellcome Fund/Howard Hughes Medical Institute scientific management course.

Workshops on time management, communication and project management were highly interactive. Professional trainers familiar with academia taught us, for example, how to stop work from piling up, something many of us had accepted as a natural consequence of a science job.

Established scientists from all over the world discussed ways of setting up collaborations, hiring personnel, getting funded and balancing work with family. Managerial skills are indispensable.

A talk on careers outside academia, such as industry and publishing, noted the transferable skills and qualities postdocs have, such as problem-solving, creativity and perseverance.

Participants gave five-minute talks in groups of ten to explain their research and to formulate five-year research and career goals. Discussing our goals in

small peer groups made us realize that we are responsible for our future and should direct it.

During a forum — the only session that included institute directors and principal investigators — we discussed many aspects of scientific careers. A key topic was Europe's rigid yet unclear academic path, which hampers funding and independence for young investigators. We proposed more peer-reviewed funding for young independent investigators and external review committees for faculty members to create better opportunities for postdocs.

The retreat combined an enthusiastic atmosphere with the comfort of confidentiality. Afterwards, some of us initiated important career moves, starting collaborations and looking for niche projects. Others began looking for a career outside academia. We all looked at our projects from a different perspective and learnt that we must invest now in our future to create the career options we need. We encourage postdocs to look at our website and be inspired to organize their own retreat.

Erik van Beers of the Netherlands Cancer Institute in Amsterdam, Anke Klerkx of the University of Amsterdam and Andrea Thiele of the Hubrecht Laboratory in Utrecht were on the retreat organizing committee.
www.mekentosj.com/postdocs

GRADUATE JOURNAL

Endurance test

For the past three months, I have spent a lot of my free time preparing for the Boston half-marathon. Training for and running the 21-kilometre race last month made me realize that getting a PhD is an endurance event.

Accomplishing your goal, whether it's finishing a race or a PhD, takes a lot of work and time. Unfortunately, all the work can be derailed by bad weather, an injury, an unlucky collaboration or an experiment gone wrong.

Company can also make all the difference. Having someone to run with gives me that extra boost. I shudder to think what graduate school would have been like without the support of classmates, friends and my boyfriend.

And sometimes you just have to stop. During the race, I stopped at water stations and walked while I drank. In those ten seconds, I caught my breath and renewed my will to run. Although I might have shaved off time if I hadn't stopped, I'm pretty sure I wouldn't have been able to run as hard. Similarly, there were many times in graduate school when I had to take mental breaks.

In the final mile of the race, I was very tired and my left leg was cramping, but I kept running because I knew I was close to the finish line. That's how I feel now about graduate school. It's been a long time since I started and there have been some tough moments, but I can only believe the end is near. Even if my pace is slower than I was hoping for, I have to keep going.

Anne Margaret Lee is at Harvard University, Boston, Massachusetts.

Shopping

Place your bets and buy into the future.

Scott Seller-Mason

Sam walked into his living room, checked his watch and switched on his computer. An image of a packed football stadium filled the screen and a voice boomed out over the roar of an expectant crowd. "And now on Channel 11, the European Football Championship final, bought to you by the Institute of Physics! Where would you be tonight without physics? And where could you be tomorrow? Support physics and support fun!"

The physics reference caused Sam's interest in the football to falter, as he considered whether he would rather go science shopping instead of watching the game. Sam had seen a research project advertised that he wanted to support, and there was one that he needed to reject. And, to be honest, he didn't actually like either of the football teams playing.

"Whaddaya reckon, Virgil? A little shopping?" Sam asked his cat, which had sauntered into the room behind him. Virgil looked at the pictures of the football game and yawned. With a wink of an eye, Sam switched his computer to Internet mode.

Along with the rest of the British public, Sam had been exposed to frantic advertising campaigns from every scientific community for the past three months. Now the time had come for the public to decide in the week-long People's Internet Vote on how taxpayers' money should be spent on science research.

The policy had stemmed from public outrage back in the early '20s over the financial demands made by scientists who spent money on spectacularly flawed research projects — the doomed zepto-technology initiatives, for example. This was the third time that people had gone to the polls in 12 years and, each time, they had become obsessed with science shopping. People chatted in cafés and argued in bars about their favourite and most hated projects, and the scientists had become commercially savvy in their attempts to woo the voters. The bookmakers did a roaring trade.

Sam flicked to one of the proposals on the government website. "Smart Fabrics! Sustainable shopping!" He had seen this on a large billboard in town. "One fabric: a million patterns, a million styles. Clothes

can be cut, shaped, coloured and styled, and then changed back to an original 'base' fabric once the design begins to date." Sam had placed a bet on this project being funded and had persuaded all his friends to vote for it, but had not yet cast his own crucial vote.

He flicked to the scientific blurb. "The fabric itself will be a thin, highly flexible photonic crystal fabricated from a biomimetic cotton, with resealable seams along facets of the crystal. Colour and pattern can be incorporated by expanding or contracting tiny spaces in the material using an entirely reversible process. The shape memory will come from a crosslinking chemical in the material, which preserves the desired shape and then decouples..."

He gave up on the technical stuff. The scientists proposed that new designs could be implemented in one of a chain of shaping and colouring salons. Fantastic. "That's a winner if ever I saw one, and most importantly, at 10:1 it will make me a few quid too," Sam said to Virgil who had sidled up to the sofa. Off Sam's vote went, via the biometric modulator, into the ether.

Sam then flicked over to one of the climate-change proposals, "The impact of the expanding hydrogen economy on climate". He swivelled his head and gazed

out of the window at his new Ford Stratos on the driveway. Sam knew that earlier climate-change projects had been incredibly influential in shifting public and government opinion against hydrocarbon use. But Sam was very proud of his new (and rather expensive) car, with the latest Metalorga hydrogen tank. He hesitated for a second, and then went for the reject option, reassuring himself that he would support the project the next time round. Maybe I'll get a few extra years out of the car before it has to be scrapped, he thought to himself.

"That's the two most important ones sorted anyway," Sam sighed, and switched back to the football. As he started to settle down, a small voice came from the direction of his cat. A faint signal from Virgil's brain had been picked up by the small antenna in Virgil's collar, which initiated a recorded voice emanating from a speaker also on the collar. "Virgil is hungry. Virgil needs feeding." Sam's heart sank. Virgil had been so quiet, he was hoping the mechanism on the collar had broken.

"Talking bloody cats," he grimaced at Virgil. "Now that is something that I really should have rejected." He headed out to the kitchen to hunt for the cat food. ■

Scott Seller-Mason is a writer living in Walla Walla, Washington, USA.

